

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2001, 13:29:06 ; Search time 8.36 Seconds

(without alignments)
114.731 Million cell updates/sec

Title: US-09-251-133-6

Perfect score: 141
Sequence: 1 RTCVLYGLHTVPEFIESQLGLLSPVSL 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Archived: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	34.8	236	1 BIL_RAT	P55062 ratu
2	49	34.8	570	1 ILVD_LACLA	002139 lactococcus
3	47	33.3	302	1 CFA2_MYCTU	011196 mycobacteri
4	47	33.3	1323	1 ADRI_YEAST	P07248 saccharomyc
5	46	32.6	183	1 LIPO_BUPMA	001584 bufo marinu
6	46	32.6	491	1 IRE3_CHICK	090643 gallus gall
7	46	32.6	467	1 PLE1_RAT	P30427 ratu
8	45	31.9	492	1 SECY_CYAPA	P25012 cyano
9	45	31.9	1400	1 RON_HUMAN	004912 homo sapien
10	44	31.2	237	1 BIL_HUMAN	P55061 homo sapien
11	44	31.2	372	1 NTUM_HIRSA	099824 rhipicephal
12	44	31.2	313	1 CDK9_HUMAN	P50750 homo sapien
13	44	31.2	593	1 SUMT_YEAST	P36130 saccharomyc
14	44	31.2	674	1 PTGA_BRELA	045288 breviabacter
15	43.5	30.9	649	1 METX_YEAST	004533 saccharomyc
16	43	30.5	111	1 YH12_YEAST	P38895 saccharomyc
17	43	30.5	229	1 Y790_METUA	058200 methanococc
18	43	30.5	223	1 YMDA_CHLAU	043826 chloroflexu
19	43	30.5	326	1 TMOF_PSEME	043826 chloroflexu
20	43	30.5	435	1 YYS3_CABEL	P49191 caenorhabdi
21	43	30.5	437	1 CCA_ARCEU	028126 archaeoglob
22	43	30.5	1131	1 DNBI_HSV7J	P55722 thizobium s
23	42	29.8	272	1 Y4YN_RHISN	P55722 thizobium s
24	42	29.8	360	1 RTCA_AOUAE	066884 aquifex aeo
25	42	29.8	434	1 FLIT_HELPJ	092133 helicobacte
26	42	29.8	434	1 FLIT_HELPJ	092133 helicobacte
27	42	29.8	524	1 MVIN_SALTY	P37169 salmonella
28	42	29.8	662	1 SL51_RABIT	P11730 corycolagus
29	42	29.8	956	1 HELI_HCMVA	P16736 human cytom
30	42	29.8	1403	1 BIRA_MOUSE	O9qws5 mus musculu
31	41	29.1	82	1 V082_ASPB7	P18557 african swi
32	41	29.1	118	1 V18_ASPB7	P18556 african swi
33	41	29.1	136	1 AT91_BOVIN	P32876 bos taurus

34	41	29.1	136	1 AT91_SHEEP	P17605 ovls arles
35	41	29.1	213	1 RISA_ECOLI	P29015 escherichia
36	41	29.1	214	1 DYA_MYCTU	P31500 mycobacteri
37	41	29.1	290	1 PROB_LEPIN	P94871 leptospira
38	41	29.1	423	1 AIM1_MOUSE	P35585 m adaptor-r
39	41	29.1	511	1 MVIN_MOUSE	P75932 escherichia
40	41	29.1	884	1 RPOB_NPYOP	O12934 orgyia pseu
41	41	29.1	893	1 PMTX_SCHPO	O13898 schizosacch
42	41	29.1	1118	1 YIP1_YEAST	P40456 saccharomyc
43	40.5	28.7	329	1 GALE_STRLI	P13226 streptomyce
44	40.5	28.7	348	1 EXOB_AZOB	O59083 azospirillum
45	40.5	28.7	391	1 BRB2_HUMAN	P30411 homo sapien

ALIGNMENTS

RESULT	ID	Score	Match	Length	Description
1	BIL_RAT	49	34.8	236	P55062 ratu
AC	P55062; 064712;				
DT	01-OCT-1996 (Rel. 34, Created)				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	01-OCT-2000 (Rel. 40, Last annotation update)				
DE	BAX INHIBITOR-1 (BI-1) (TESTIS ENHANCED GENE TRANSCRIPT).				
GN	BIL OR TEGT.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_Taxid=10116;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=SPRAGUE-DAWLEY; TISSUE=Testis;				
RX	MEDLINE=94281747; PubMed=8012111;				
RA	Walter L., Dicks B., Rothenmel E., Heyens M., Szpieler C., Leyvan G.,				
RA	Guenther E.;				
RT	"A novel, conserved gene of the rat that is developmentally regulated				
RL	in the testis."				
CC	Mamm. Genome 5:216-221(1994).				
CC	-1- FUNCTION: SUPPRESSOR OF APOPTOSIS (BY SIMILARITY).				
CC	-1- SUBUNIT: INTERACTS WITH BCL2 AND BCL-XL (BY SIMILARITY).				
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).				
CC	-1- TISSUE SPECIFICITY: HIGHLY ABUNDANT IN TESTIS.				
CC	-1- SIMILARITY: BELONGS TO THE BIL FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
DR	EMBL; X75855; CA53470.1; -				
DR	EMBL; X75856; CA53471.1; -				
DR	InterPro: IPR002199; -				
DR	Pfam: PF01027; UPR0005; 1.				
DR	PROSITE; PS01243; BIL; 1.				
KW	Apoptosis; Transmembrane.				
FT	TRANSMEM 30				POTENTIAL.
FT	TRANSMEM 53				POTENTIAL.
FT	TRANSMEM 88				POTENTIAL.
FT	TRANSMEM 112				POTENTIAL.
FT	TRANSMEM 139				POTENTIAL.
FT	TRANSMEM 159				POTENTIAL.
FT	TRANSMEM 166				POTENTIAL.
FT	TRANSMEM 186				POTENTIAL.
FT	TRANSMEM 206				POTENTIAL.
FT	TRANSMEM 226				POTENTIAL.
SQ	SEQUENCE 236 AA; 26358 MW; FFA12EC1DC87537 CRC64;				

Query Match 34.8%; Score 49; DB 1; Length 236;
Best local similarity 40.9%; Pred. No. 2.4;
Matches 9; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2001, 13:25:16 ; Search time 15.4 Seconds
(without alignments)
138.499 Million cell updates/sec

Title: US-09-251-133-6

Perfect score: 141
Sequence: 1 RTCVLGYLHIVPEFIESQLGLSPVSL 28

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Indexed: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR_68:
2: PIR1:
3: PIR2:
4: PIR3:
5: PIR4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	37.6	931	2 H86387	hypothetical prote
2	50	35.5	813	2 A72203	cellobiose-phospho
3	49	34.8	236	2 S42069	TEGT protein - rat
4	49	34.8	275	2 A56641	probable membrane
5	49	34.8	570	2 S53137	probable phosphogl
6	49	34.8	570	2 G86777	dihydroxy-acid deh
7	48	34.0	702	2 E72775	probable helicase
8	47	33.3	302	2 B70746	probable cma2 pro
9	47	33.3	308	2 S72886	hypothetical prote
10	47	33.3	532	2 C69067	phenylalanine-trn
11	47	33.3	656	2 B86033	hypothetical prote
12	47	33.3	1333	2 A24534	regulatory protein
13	46.5	33.0	554	2 T18858	hypothetical prote
14	46	32.6	159	2 T08974	hypothetical prote
15	46	32.6	179	2 H71367	conserved hypochet
16	46	32.6	183	2 S25465	lipocalin - giant
17	46	32.6	491	2 S56753	interferon regulat
18	46	32.6	662	2 D86320	hypothetical prote
19	46	32.6	4687	1 A39638	plectin - rat
20	45.5	32.3	616	2 T00894	hypothetical prote
21	45.5	32.3	643	2 D86167	protein F21B7.27
22	45.5	32.3	871	2 H72957	hypothetical prote
23	45.5	32.3	982	2 S00954	pol polyprotein -
24	45.5	32.3	2048	2 C84609	hypothetical prote
25	45	31.9	310	2 T30674	hypothetical prote
26	45	31.9	388	2 F70430	hypothetical prote
27	45	31.9	432	2 E96880	Na+/H+ antiporter
28	45	31.9	477	2 D82390	preprotein translo
29	45	31.9	492	2 T06875	

30	45	31.9	1319	2 H84542	hypothetical prote
31	45	31.9	1400	1 T38185	protein-tyrosine k
32	44.5	31.6	240	2 E69004	hypothetical prote
33	44.5	31.6	648	2 T23864	hypothetical prote
34	44	31.2	237	2 T38334	TEGT (testis enhan
35	44	31.2	313	2 T11160	NADH dehydrogenase
36	44	31.2	317	2 C83652	hypothetical prote
37	44	31.2	344	2 H71825	hypothetical prote
38	44	31.2	372	2 A55262	probable histidine
39	44	31.2	593	2 S38145	protein kinase (PC
40	44	31.2	612	2 T05331	uroporhyrinogen m
41	44	31.2	648	2 T41915	hypothetical prote
42	43.5	30.9	420	2 F69144	O-antigen transpor
43	43.5	30.9	649	2 S43644	hypothetical prote
44	43.5	30.9	650	2 T38692	probable serine/th
45	43.5	30.9	1169	2 T30207	dynein heavy chain

ALIGNMENTS

RESULT 1
H86387
hypothetical protein AAG29216.1 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H86387
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, N.F.; Hughes, B.; Huizier, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzla, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; M01D:21016719
A:Accession: H86387
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1931 <STO>
A:Cross-references: GB:AE005172; NID:g11079505; PIDN:AAG29216.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 37.6% Score 53; DB 2; Length 931;
Best local Similarity 37.5% Pred. No. 6.5;
Matches 9; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

OY 3 CVLGYLHIVPEFIESQLGLSPV 26
DB 246 CLNCFVONFPLIESELMGMFSL 269

RESULT 2
A72203
cellobiose-phosphorylase - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: A72203
A:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Blic, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, C.M.
Nature 399, 323-329, 1999
A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; M01D:99287316
A:Accession: A72203
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-813 <MAN>
A:Cross-references: GB:AE001822; GB:AE000512; NID:g4982429; PIDN:AAD36910.1; PID:g498

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2001, 13:22:45 ; Search time 19.06 Seconds
(without alignments)
89.059 Million cell updates/sec

Title: US-09-251-133-6

Perfect score: 141
Sequence: 1 RTCVLGYLHYVEPIEISQLGLSPVSL 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Filtered: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

A-Geneseq_0601:*

1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT:*

2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT:*

3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT:*

4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT:*

5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT:*

6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT:*

7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT:*

8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT:*

9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT:*

10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT:*

11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT:*

12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT:*

13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT:*

14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT:*

15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT:*

16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:*

17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT:*

18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT:*

19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT:*

20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:*

21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:*

22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51	36.2	552	22	AA848241
2	49.5	35.1	188	19	AAW76633
3	49	34.8	236	22	AA87615
4	49	34.8	570	15	AA854219
5	47.5	33.7	171	18	AAV11333
6	47.5	33.7	173	19	AAW38569
7	46.5	33.0	173	19	AAW80667
8	46	32.6	127	21	AA824844
9	46	32.6	169	21	AA824843
10	46	32.6	181	21	AA824842
11	46	32.6	183	22	AA890659

12	45.5	32.3	125	18	AAW28030	Staphylococcus aur
13	45	31.9	286	21	AA824612	Arabidopsis thalia
14	45	31.9	328	21	AA824611	Arabidopsis thalia
15	45	31.9	354	21	AA824610	Arabidopsis thalia
16	45	31.9	822	20	AAW73486	Cellobiose phospho
17	45	31.9	1400	20	AAW82791	Human RON receptor
18	44	31.2	122	21	AA803753	Human secreted pro
19	44	31.2	139	22	AA861120	Human herpesvirus
20	44	31.2	237	19	AAW73136	Bax inhibitor Bi-1
21	44	31.2	255	21	AA858178	Lung cancer associ
22	44	31.2	372	17	AAW04869	Cyclin-dependent k
23	44	31.2	372	17	AAW04871	Phosphorylation de
24	44	31.2	374	21	AA858375	Lung cancer associ
25	44	31.2	648	20	AAW96198	Human herpesvirus-
26	44	31.2	648	20	AAW96199	Human herpesvirus-
27	44	31.2	2595	20	AAW39297	Spn a polyketide
28	43.5	30.9	160	21	AA829660	Human membrane-ass
29	43.5	30.9	1098	22	AA876533	Corynebacterium gl
30	43.5	30.9	1874	22	AA876532	Corynebacterium gl
31	43	30.5	286	20	AAW73538	Chlamydia pneumoni
32	43	30.5	308	22	AA836408	Secreted protein K
33	43	30.5	326	22	AA860228	Pseudomonas mendoc
34	43	30.5	378	22	AA865658	Novel protein kina
35	43	30.5	435	21	AA800189	Putative polyunsat
36	43	30.5	501	21	AA854157	Human pancreatic c
37	43	30.5	535	19	AAW77299	Amino acid sequenc
38	43	30.5	535	22	AA865657	Novel protein kina
39	42.5	30.1	372	22	AAV72816	Calendula officina
40	42.5	30.1	965	21	AA842561	Human ORFX ORF2325
41	42	29.8	69	21	AA844200	Human cancer assoc
42	42	29.8	259	20	AAV35625	Chlamydia pneumoni
43	42	29.8	269	20	AAV08563	B. subtilis hydrol
44	42	29.8	394	19	AAW49681	Open reading frame
45	42	29.8	459	21	AA824717	Arabidopsis thalia

ALIGNMENTS

RESULT 1	AA848241	standard; Protein: 552 AA.
ID	AA848241;	
AC	AA848241;	
XX	02-Apr-2001 (first entry)	
DE	Amino acid sequence of bZIP2 ORF1 protein.	
XX		
KW	Transcription factor; seed storage protein; lectin; oil-body protein;	
KW	Pv-Seed factor-1; ROM1; Vicilin-box binding protein-1; ROM2; 7S-globulin;	
KW	phaseolin; PNA-L; bean; nuclear protein; promoter; ORF; bZIP;	
KW	basic leucine zipper.	
XX		
OS	Phaseolus vulgaris.	
XX		
FT	Key	Location/Qualifiers
FT	Misc-difference 1..552	/note- "Xaa are residues encoded by internal stop codons"
XX		
PN	US6160202-A.	
XX		
PD	12-DEC-2000.	
XX		
PF	06-FEB-1997;	97US-0796899.
XX		
PR	07-OCT-1994;	94US-0319544.
XX		
RA	(UYMA-) UNIV MARYLAND BALTIMORE COUNTY.	
XX		
PI	Chern M, Bustos MM;	
XX		
DR	WPI; 2001-079619/09.	

This Page Blank (uspto)

DR N-PSDB; AAC84565.

XX Novel transcription factor gene which encodes transcription factor
PT protein that targets promoters of genes encoding seed storage proteins
PT are useful for modulating seed storage protein expression in dicot seed
PT crops

XX Disclosure; Columns 31-36; 67pp; English.

XX The invention relates to an isolated transcription factor gene which is
CC expressed in a recombinant maturing dicot seed and which encodes a
CC transcription factor protein which targets a promoter of a gene encoding
CC seed storage proteins, lectins or oil-body proteins. The transcription
CC factors isolated are Pv-Seed Factor-1 (ROM1) and Vicilin-box binding
CC protein-1 (ROM2). These factors bind to 7S-globulin (b-phaseolin) or
CC lectin (PHA-L) promoters. The transcription factor gene is useful for
CC enhancing or reducing expression of seed storage protein, lectin or
CC oil-protein genes in dicot seed crops. The present sequence represents
CC the amino acid sequence of bz1p2 (basic leucine zipper) ORF1 protein.

XX Sequence 552 AA;

Query Match 36.2%; Score 51; DB 22; Length 552;
Best Local Similarity 55.0%; Pred. No. 6.7;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 4 VLGYLHVPEFIESQLGL 23
DB 121 11ghvhtvkefph111g1 140

RESULT 2

AAW76633
ID AAW76633 standard; Protein; 188 AA.

AC AAW76633;

DT 12-JUL-1999 (first entry)

DE Human herpes virus type 8 FLIP ORF 71 protein.

KW Death effector domain; human; murine; anti-apoptotic; treatment;
KM HIV infection; autoimmune disease.

XX Human herpes virus.

DE19713393-A1.

XX 08-OCT-1998.

PF 01-APR-1997; 97DE-1013393.

PR 01-APR-1997; 97DE-1013393.

PA (TSCCH/) TSCHOPP J.
PA (APOT-) APOTTECH SA.

PI Bodmer J, Burns K, French EL, Hahne M, Hoffmann K;
PI Irmeler M, Rimoldi D, Schneider P, Schroeter M, Steiner V;
PI Thome M, Tschoep J, Hofmann K;

DR WPI; 1998-532710/46.

XX New DNA encoding for anti-apoptotic gene product - used to treat HIV
PT infections and autoimmune diseases

XX Claim 16; Fig 16; 45pp; German.

CC This invention describes novel human and mouse anti-apoptotic gene
CC products which contain at least one death effector domain. The products
CC of the invention are used in the treatment of HIV infections and
CC autoimmune diseases. This sequence represents the human Herpes virus

CC type 8 ORF 71 FLIP protein which is used in the method of the invention.

XX Sequence 188 AA;

Query Match 35.1%; Score 49.5; DB 19; Length 188;
Best Local Similarity 44.0%; Pred. No. 3.4;
Matches 11; Conservative 4; Mismatches 9; Indels 1; Gaps 1;

OY 1 RTCVL-GYLHVPEFIESQLGLS 24
DB 65 rrcilrllhndpflernhlagms 89

RESULT 3

AAB87615
ID AAB87615 standard; protein; 236 AA.

AC AAB87615;

DT 15-MAY-2001 (first entry)

DE Bovine mammary tissue derived protein #6.

KW Bovine; mammary gland; cancer; tumour; angiogenesis.

OS Bos taurus.

PN WO200114553-A1.

PD 01-MAR-2001.

PF 23-AUG-2000; 2000MO-NZ00166.

PR 23-AUG-1999; 99US-0150330.

PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (NZPA-) NEW ZEALAND PASTORAL AGRIC RES INST LTD.

PI Havukkala IJ, Gleen M, Grigor MR, Molenaar AJ;

DR WPI; 2001-226619/23.

XX New polypeptides and polynucleotides encoding the polypeptides, which
PT are expressed in bovine mammary gland tissue, useful for stimulating
PT mammary gland growth or function, or inducing differentiation of milk
PT producing cells

XX Claim 11; Page 62; 97pp; English.

PS The present invention relates to proteins derived from bovine

CC mammary gland cells. The invention is useful for stimulating the

CC bovine mammary gland cell growth and function, inhibiting the

CC growth of various mammary gland cancer cells, inhibiting

CC angiogenesis and vascularization of tumours, or modulating

CC the growth of blood vessels in a mammal.

XX Sequence 236 AA;

Query Match 34.8%; Score 49; DB 22; Length 236;
Best Local Similarity 40.3%; Pred. No. 5.3;
Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

OY 7 VLHVPEFIESQLGLSPVL 28
DB 46 ylhvthfigqllsalslgl 67

RESULT 4

AAR54219
ID AAR54219 standard; Protein; 570 AA.

XX

Two Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2001, 13:28:41 ; Search time 23.02 Seconds
(without alignments)
160.927 Million cell updates/sec

Title: US-09-251-133-6
Perfect score: 141
Sequence: 1 RCVYGLHYIPEETESQLGLGLSPVSL 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP:REMBL_16:*
2: SP:Archea:*
3: SP:Bacteria:*
4: SP:Fungi:*
5: SP:Human:*
6: SP:Invertebrate:*
7: SP:Mammal:*
8: SP:Organelle:*
9: SP:Phage:*
10: SP:Plant:*
11: SP:Rodent:*
12: SP:Unclassified:*
13: SP:Vertebrate:*
14: SP:Virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	36.9	239	8	09MS96
2	52	36.9	688	9	09VXB7
3	50	35.5	812	2	052504
4	50	35.5	813	2	087964
5	50	35.5	813	2	09X2G3
6	49	34.8	275	2	006111
7	48.5	34.4	469	5	09VFE9
8	48	34.0	99	6	09S309
9	48	34.0	512	5	09VZU3
10	48	34.0	626	5	09U3X4
11	48	34.0	702	1	09YFQ8
12	47	33.3	146	10	09LRT3
13	47	33.3	308	2	049807
14	47	33.3	496	3	004919
15	47	33.3	532	1	027545
16	46.5	33.0	415	6	018856
17	46.5	33.0	550	10	09LTM4
18	46.5	33.0	554	5	09J174
19	46.5	33.0	681	2	09L1X1

20	46.5	33.0	913	5	097272	097272 Plasmodium
21	46	32.6	169	10	09SZV1	09szvl arabidopsis
22	46	32.6	179	2	083125	083125 treponema p
23	46	32.6	662	10	09M978	09m978 arabidopsis
24	46	32.6	964	11	09QX51	09qx51 mus musculu
25	46	32.6	4684	4	015149	015149 homo sapien
26	45.5	32.3	395	6	018857	018857 phascogale
27	45.5	32.3	425	6	018859	018859 droncllops
28	45.5	32.3	425	6	018860	018860 vomatus ur
29	45.5	32.3	643	10	09LRS9	09lrs9 arabidopsis
30	45.5	32.3	871	1	09YCK9	09yck9 aeropyrum p
31	45.5	32.3	2048	10	09SIF1	09sif1 arabidopsis
32	45	31.9	288	2	096312	096312 burholderi
33	45	31.9	310	14	098240	098240 molluscum c
34	45	31.9	388	2	067472	067472 aquifex aeo
35	45	31.9	405	5	09U3D1	09u3d1 caenorhabd1
36	45	31.9	409	5	091783	091783 polyorchis
37	45	31.9	428	2	09EXG4	09exg4 vibrio chol
38	45	31.9	432	10	004473	004473 arabidopsis
39	45	31.9	477	2	09KRT5	09krt5 vibrio chol
40	45	31.9	480	2	09E8X3	09e8x3 pseudomonas
41	45	31.9	558	3	087161	087161 aspergillus
42	45	31.9	811	2	059316	059316 clostridium
43	45	31.9	822	2	066264	066264 cellvibrio
44	45	31.9	1013	14	09Q078	09q078 chimpanzee
45	45	31.9	1319	10	09SLE9	09sle9 arabidopsis

ALIGNMENTS

RESULT 1
ID 09MS96 PRELIMINARY: PRT: 239 AA.
AC 09MS96;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 26.8 KDA PROTEIN.
OS Galdieria sulphuraria.
OC Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
OC Galdieria.
OX NCBI_TaxID=130081;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-UTEX 2393;
RA Whitney S.M., Andrews J.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
EMBL: AF233069; AAF81686.1; -
DR InterPro: IPR001687; -
DR InterPro: IPR003439; -
DR InterPro: IPR003593; -
DR Pfam: PF00005; ABC_tran.1.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR SMART: SM00382; AAA; 1.
KW Hypothetical protein; Chloroplast.
SQ SEQUENCE 239 AA: 26836 MW: 58A2868B54A85FD CRC64;

Query Match 36.9%; Score 52; DB 8; Length 239;
Best Local Similarity 44.0%; Pred. No. 7.1;
Matches 11; Conservative 5; Mismatches 7; Indels 2; Gaps 1;
OY 1 RCVYGLHYIPEETESQLGLSP 25
DB 14 RCIIDNLTMT--FSQYELGLSP 36
RESULT 2
OYVXB7
ID 09VXB7 PRELIMINARY: PRT: 688 AA.
AC 09VXB7;

mis Page Blank (uspto)


```

seq_name: gb_gss23:A2238007
seq_documentation_block:
LOCUS      A2238007             324 bp    DNA                GSS           14-JUN-2000
DEFINITION RPCI-23-72B19.TV RPCI-23 Mus musculus genomic clone RPCI-23-72B19,
ACCESSION   A2238007
VERSION     A2238007.1 GI:8546053
KEYWORDS
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 324)
AUTHORS     Zhao,S., Nieman,W., Feldblum,T., Malek,J., Shatsman,S., Akintet
            ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
            Mouse BAC End Sequences from Library RPCI-23
            Unpublished (1999)
COMMENT     Other GSSs: RPCI-23-72B19.TJ
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel.: 301 838 0200
            Fax: 301 838 0208
            Email: szhaoc@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@edjong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html
            Plate: 72 row: B column: 19
            Seq primer: T7
            Class: BAC ends.
FEATURES             location/qualifiers
     source          1..324
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-72B19"
                     /clone_id="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain. Vector: pBACE3.6; Site.1:
                     EcoRI; Site.2: EcoRI; Female C57Bl/6J mouse kidney and/or
                     brain genomic DNA was isolated and partially digested
                     with a combination of EcoRI and EcoRI Methylase. Size
                     selected DNA was cloned into the pBACE3.6 vector at the
                     EcoRI sites. The ligation products were transformed into
                     DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT        97 a         57 c         56 g         114 t
ORIGIN
alignment_scores:
Quality:          61.50              Length:          28
Ratio:            2.929              Gaps:            1
Percent Similarity: 75.000          Percent Identity: 46.429
alignment_block:
US-09-251-133-6 x A2238007 ..
Align seg 1/1 to: A2238007 from: 1 to: 324
4 ValLeUGLYTYrLeuHisIleValProGluPheIleGluSerGlnLeu 20
:::||||| |||||:::|||||:::|||||:::|||||:::|||||:::
147 TTGCCTTCATTATTCATGTCACTGAATTTATAGATGTAAGTAACTGAT 196

```

```

20 u.....GlyleUserProValSerHeu 28
      :      :::  |||::||||||
197 TATTACTGTCTGCAATCTTCTCTATCCCTG 230

seq_name: gb_est86:BF341640

seq_documentation_block:
LOCUS      BF341640          814 bp      mRNA          EST          22-NOV-2000
DEFINITION      602016073F1 NCL_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4151706
              5', mRNA sequence.
ACCESSION      BF341640
VERSION        BF341640.1  GI:11288136
KEYWORDS
SOURCE
human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 814)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL      National Institutes of Health, Mammalian Gene Collection (MGC)
              Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaps@email.nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM9417 row: a column: 19
High quality sequence start: 8
High quality sequence stop: 707.
Location/Qualifiers
1. 814
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /clone="IMAGE:4151706"
   /clone_id="NCL_CGAP_Brn64"
   /tissue_type="gliblastoma with EGFR amplification"
   /lab_host="DH10B (T1 phage-resistant)"
   /note="Organ: Brain; Vector: PCMV-SPORT6; Site_1: NotI;
   Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
   Average insert size 1.57 kb. Constructed by Life
   Technologies. Note: this is a NCL_CGAP Library."
BASE COUNT      236 a      155 c      162 g      261 t
ORIGIN
alignment_scores:
      Quality:      61.00      Length:      22
      Ratio:      3.588      Gaps:      0
Percent Similarity: 77.273      Percent Identity: 54.545

alignment_block:
US-09-251-133-6 x BF341640      ..

Align seg 1/1 to: BF341640 from: 1 to: 814

6 G1YTYTLEUHSITLVALPBGGLUPHETLEUGSERGLNLEUHGJLYLE 22
      |||||      |||::|||:::      ||||||      |||
101 GGCTATTCATCATATTCCTCCCAATTCCTTCTTTCAGCCAACTCAAGGATA 550

22 uleuserProValSer 27
      :::|||::|||:::
551 TGTATCACCTTGACT 566

seq_name: gb_gss25:AZ366689

seq_documentation_block:
LOCUS      AZ366689          699 bp      DNA          GSS          02-OCT-2000
DEFINITION      IM0116102F Mouse 10kb plasmid U06C1M library Mus musculus genomic

```

```

25 CGTCAGACAGTTATGAGGCTGCTGCATTTGGAGCCATTCCTCCTCAAGCT 74
11 .....ValProGluPheIleGluSerIleuLeuGlyLeuLeuSerP 25
       |||
75 GACTGAACACCGGAGCTTATGCTTCTCCCGCTTTGGCGCTCTGCTCT 124
25 roVal 26
   |||
125 CAGTC 129

seq_name: gb_est52:AW859667

seq_documentation_block:
LOCUS      AW859667              499 bp      mRNA          EST          19-MAY-2000
DEFINITION RC4-CT0360-141299-011-d04 CT0360 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW859667
VERSION     AW859667.1  GI:7955360
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 499)
Dias Neto,E., Garcia Correa,R., Veijovski-Almeida,S., Briones,M.R.,
Negal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=6t2-RC4-CT0360-141
299-011-d04&t3=1999-12-14&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 22
Location/Qualifiers
1..499
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CF0360"
/der_stage="Adult"
/note="Organ: colon; Vector: puc18; Site.1: Sma1; Site.2:
Sma1; A mini-library was made by cloning products derived
from ORESYES PCR (U.S. Letters Patent application No. 199
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT      172 a      132 g      102 t
ORIGIN
alignment_scores:
Quality:      58.00      Length:      28
Ratio:      3.053      Gaps:      1
Percent Similarity: 67.857      Percent Identity: 50.000

alignment_block:
US-09-251-133-6 x AW859667 ..

```

Align seg 1/1 to: AW859667 from: 1 to: 499

```

1  ArgThrcysValLeuGlyTyrLeuHisIleValProGluPheIleGluSer 17
   ::::::::::::::::::::
324 CAGACATTT.....GGCAGATGCCATATAGTCCACAGCTTCTCGAGAGC 367
   |||
17  rGlnLeuLeuGlyLeuLeuSerProValSerLeu 28
   ::|
368 TCAGGCTGCAGCATTTGCTTGACCCAGCACTTG 401
   |||

```

seq_name: gb_gss18:AO880197

seq_documentation_block:

LOCUS AO880197 487 bp DNA GSS 09-NOV-1999
 DEFINITION HS-5037.B1.G03.T7.RPCI-11 Human Male BAC Library Homo sapiens
 genomic clone Plate=8805 COL=5 Row=N, DNA sequence.
 ACCESSION AO880197
 VERSION AO880197.1 GI:6311664
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 487)
 Mahaltras,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc.Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

JOURNAL MEDLINE
 COMMENT Contact: Mahaltras Gg, Wallace Jc, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pletier de Jong
 (pleterdejong@med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
 or from Research Genetics (<http://www.htsc.washington.edu>)
 Plate: 8805 row: N column: 5
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 487.

FEATURES

source
 Location/Qualifiers

```

1..487
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone_lib="Plate=8805 COL=5 Row=N"
  /clone_lib="RPCI-11 Human Male BAC Library"
  /sex="male"
  /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"
BASE COUNT 159 a 83 c 100 g 138 t 7 others
ORIGIN

```

alignment_scores:
 Quality: 57.50 Length: 24
 Ratio: 2.875 Gaps: 1
 Percent Similarity: 83.333 Percent Identity: 45.833

alignment_block:

US-09-251-133-6 x AO880197 ..

Align seg 1/1 to: AO880197 from: 1 to: 487

```

2  ThrCysValLeuGlyTyrLeuHisIleValProGluPheIleGluSer 18
   ::::::::::::::::::::
253 ACATGATATATTA...TATATATATATAGTCCACAGCACTTTAGAGCGTGA 299
   ::|
18  nLeuLeuGlyLeuLeuSerPro 25
   ::|
300 GGTGGAGCATCACTTGACCA 321
   |||

```

seq_name: gb_est95:BG004824

seq_documentation_block:

LOCUS BG004824 344 bp mRNA EST 24-JAN-2001
 DEFINITION RC5-GN0132-211100-023-A04 GN0132 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BG004824
 VERSION BG004824.1 GI:12446375
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 344)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 JOURNAL MEDLINE
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC5&t2=RC5-GN0132-211100-023-A04&t3=2000-11-21&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence start: 59
 High quality sequence stop: 340.

FEATURES

source
 Location/Qualifiers

```

1..344
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone_lib="GN0132"
  /dev_stage="Adult"
  /note="Organ: Placenta_normal; Vector: puc18; Site_1: SmaI
; Site_2: SmaI; A mini-library was made by cloning
products derived from ORESTRS PCR (U.S. Letters Patent
application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the puc18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."
BASE COUNT 94 a 83 c 91 g 76 t
ORIGIN

```

alignment_scores:
 Quality: 57.00 Length: 17
 Ratio: 3.800 Gaps: 0
 Percent Similarity: 88.235 Percent Identity: 64.706

alignment_block:

US-09-251-133-6 x BG004824/rev ..

Align seg 1/1 to reverse of: BG004824 from: 1 to: 344

Align seg 1/1 to reverse of: BE910433 from: 1 to: 818

8	LeuHisIleValProGluPheIleGluSerGlnLeuLeuGlyLeuLeuSe	24
217	CTGCATATTAATCCGACCATATATACAGCTCCAACTACTGGGCTTAATTTC	168
24	r	24
167	C	167

This Page Blank (uspto)

OM of: US-09-251-133-6 to: Issued_Patents_NA.* out_format : pfs
 Date: Aug 7, 2001 2:26 PM

About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=framed-p2n.model -DEF=x1p
 -O/cgn2_1/USPTO.spool/US09251133/runat.06082001.175547.16513/app.query.fasta.1.84
 -DB=Issued_Patents_NA -OFM=fastap -SUFFIX=rni -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000
 -GAPOP=6.000 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -DELOP=6.000 -DEEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -TRANS=human40.cdi -LIST=45 -MATRIX=bloms62
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFM=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=200000000
 -USR=US09251133.@cgn1_1_50 -NCPV=6 -ICPV=3 -LONGLOG -NO_XLPRX
 -WAIT -THREADS=1

Search information block:

Query: US-09-251-133-6
 Query length: 28
 Database: Issued_Patents_NA.*
 Database sequences: 324599
 Database length: 94655562
 Search time (sec): 66.090000

score list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-09-176-657-5-		57.00	143.77	2.16	1506
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-348-353-16-		52.50	121.33	38.36	3744
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-08-465-965-16-		52.50	121.33	38.36	3744
/cgn2_6/ptodata/1/ina/6A.COMB.seq:US-08-465-966-16-		52.50	121.33	38.36	3744
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-666-798-1-		51.00	134.23	7.33	640
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-692-692-1-		51.00	134.23	7.33	640
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-096-071-1-		51.00	134.23	7.33	640
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-796-899-26+		51.00	124.87	24.38	1657
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-1+		49.00	133.59	7.97	372
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-1+		49.00	133.59	7.97	372
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-5-		49.00	120.62	42.04	1389
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-10-		49.00	120.62	42.15	1392
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-10-		49.00	120.62	42.15	1392
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-7-		49.00	120.25	44.07	1442
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-20-		49.00	120.10	44.92	1464
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-16-		49.00	118.79	53.17	1673
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-16-		49.00	118.79	53.17	1673
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-22-		49.00	118.05	58.44	1803
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-41-		49.00	117.39	63.60	1928
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-43-		49.00	117.39	63.60	1928
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-35-		49.00	117.39	63.60	1928
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-35-		49.00	117.39	63.60	1928
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-39-		49.00	117.39	63.60	1929
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-39-		49.00	117.38	63.68	1930
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-18-		49.00	116.65	69.95	2079
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-18-		49.00	116.65	69.95	2079
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-32-		49.00	114.88	87.84	2490
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-32-		49.00	114.88	87.84	2490
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-26-		49.00	113.12	110.01	2976
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-26-		49.00	113.12	110.01	2976
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-24-		49.00	113.03	111.28	3003
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-24-		49.00	113.03	111.28	3003
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-30-		49.00	111.27	139.46	3591
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-30-		49.00	111.27	139.46	3591
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-28-		49.00	111.20	140.79	3618
/cgn2_6/ptodata/1/ina/6B.COMB.seq:US-09-352-159-28-		49.00	111.20	140.79	3618

/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-403-866-11+ 49.00 98.82 688.53 12720
 /cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-557-309B-21- 48.00 121.51 37.49 936
 /cgn2_6/ptodata/1/ina/6A.COMB.seq:US-08-834-306-21- 48.00 121.51 37.49 936
 /cgn2_6/ptodata/1/ina/6B.COMB.seq:US-08-993-674A-21- 48.00 121.51 37.49 936
 /cgn2_6/ptodata/1/ina/5A.COMB.seq:US-07-851-976B-7- 48.00 114.32 94.27 1943

seq_name: /cgn2_6/ptodata/1/ina/6A.COMB.seq:US-09-176-657-5

seq_documentation_block:

Sequence 5, Application US/09176657
 Patent No. 6020164
 GENERAL INFORMATION:
 APPLICANT: Bandman, Olga
 APPLICANT: Tang, Y. Tom
 APPLICANT: Corley, Neil C.
 APPLICANT: Guegler, Karl J.
 APPLICANT: Lu, Alina
 APPLICANT: Baughn, Mariah R.
 TITLE OF INVENTION: HUMAN RNA BINDING PROTEINS
 FILE REFERENCE: PF-0611 US
 CURRENT APPLICATION NUMBER: US/09/176,657
 CURRENT FILING DATE: 1998-10-21
 NUMBER OF SEQ ID NOS: 9
 SOFTWARE: PERL Program
 SEQ ID NO 5
 LENGTH: 1506
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE: -
 OTHER INFORMATION: 1250374
 US-09-176-657-5

alignment_scores:

Quality: 57.00 Length: 17
 Ratio: 3.80 Gaps: 0
 Percent Similarity: 88.235 Percent Identity: 64.706

alignment_block:

US-09-251-133-6 x US-09-176-657-5/rev ..

Align seg 1/1 to reverse of: US-09-176-657-5 from: 1 to: 1506

8 LeuHStIaValProGluPheIleGluSerGlnLeuGlyLeuLeuSe 24
 |||||:::||||: ||||: |||||:::|||||:::|||||
 858 CGCATATACCTCGGACCATATACCTGAGCTTATATTC 809

24 r 24
 808 C 808

seq_name: /cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-348-353-16

seq_documentation_block:

Sequence 16, Application US/08348353
 Patent No. 5932217
 GENERAL INFORMATION:
 APPLICANT: Tuomanen, Elaine
 APPLICANT: Masure, Robert
 TITLE OF INVENTION: Antibody Recognizing Endothelial Cell
 TITLE OF INVENTION: Ligand for Leukocyte CR3
 NUMBER OF SEQUENCES: 38
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Klauber & Jackson
 STREET: 411 Hackensack Avenue
 CITY: Hackensack
 STATE: New Jersey
 COUNTRY: U.S.A.
 ZIP: 07601
 COMPUTER READABLE FORM:
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

```
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/348,353
FILING DATE: 30-NOV-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Jackson, David
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-097CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 3744 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3744
PUBLICATION INFORMATION:
AUTHORS: Delisse-Gathoye, et al.
TITLE: Cloning, Partial Sequence, Expressions, and
TITLE: Antigenic Analysis of the Filamentous
TITLE: Hemagglutinin Gene of Bordetella Pertussis
JOURNAL: Infection and Immunity
VOLUME: 58
ISSUE: 9
PAGES: 2895-2905
DATE: September-1990
US-08-348-353-16

alignment_scores:
Quality: 52.50 Length: 28
Ratio: 2.625 Gaps: 1
Percent Similarity: 71.429 Percent Identity: 46.429

alignment_block:
US-09-251-133-6 x US-08-348-353-16/rev ..

Align seg 1/1 to reverse of: US-08-348-353-16 from: 1 to: 3744
2 ThrcysvalleuglytylleuNISlEVALProglupheilegiuserG1 18
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1925 ACCTGACATCTTGCCCGCGCTGCGCGCGCGCGCGGAGTCTGACGCGTGC 1876
18 nleuenglyleuLeuser...ProvalSerleu 28
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1875 GCTCTTGCCATGACCTGCGCGCGCGCGCGTCTGCTG 1842

seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-08-465-965-16

seq_documentation_block:
; Sequence 16, Application US/08465965
; GENERAL INFORMATION:
; APPLICANT: Tuomanen, Elaine
; APPLICANT: Masure, Robert
; TITLE OF INVENTION: Antibody Recognizing Endothelial Cell
; TITLE OF INVENTION: Ligand for Leukocyte CR3
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,965
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/348,353
FILING DATE: 30-NOV-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/247,572
FILING DATE: 23-MAY-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/03725
FILING DATE: 04-MAY-1992
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/695,613
FILING DATE: 03-MAY-1991
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Jackson, David
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-097CIP1DIV2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 3744 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3744
PUBLICATION INFORMATION:
AUTHORS: Delisse-Gathoye, et al.
TITLE: Cloning, Partial Sequence, Expressions, and
TITLE: Antigenic Analysis of the Filamentous
TITLE: Hemagglutinin Gene of Bordetella Pertussis
JOURNAL: Infection and Immunity
VOLUME: 58
ISSUE: 9
PAGES: 2895-2905
DATE: September-1990
US-08-465-965-16

alignment_scores:
Quality: 52.50 Length: 28
Ratio: 2.625 Gaps: 1
Percent Similarity: 71.429 Percent Identity: 46.429

alignment_block:
US-09-251-133-6 x US-08-465-965-16/rev ..

Align seg 1/1 to reverse of: US-08-465-965-16 from: 1 to: 3744
2 ThrcysvalleuglytylleuNISlEVALProglupheilegiuserG1 18
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1925 ACCTGACATCTTGCCCGCGCTGCGCGCGCGCGGAGTCTGACGCGTGC 1876
18 nleuenglyleuLeuser...ProvalSerleu 28
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1875 GCTCTTGCCATGACCTGCGCGCGCGCGCGTCTGCTG 1842

seq_name: /cgn2_6/ptodata/1/lna/6A_COMB.seq:US-08-465-966-16

seq_documentation_block:
; Sequence 16, Application US/08465966
```

Patent NO. 6015560
GENERAL INFORMATION:
APPLICANT: Tuomanen, Elaine
APPLICANT: Masure, Robert
TITLE OF INVENTION: Antibody Recognizing Endothelial Cell
TITLE OF INVENTION: Ligand for Leukocyte CR3
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,966
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/348,353
FILING DATE: 30-NOV-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/247,572
FILING DATE: 23-MAY-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/03725
FILING DATE: 04-MAY-1992
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/695,613
FILING DATE: 03-MAY-1991
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Jackson, David
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-097CIPDIV2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-345-1684
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 3744 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3744
PUBLICATION INFORMATION:
AUTHORS: Delisse-Gathoye, et al.
TITLE: Cloning, Partial Sequence, Expressions, and
TITLE: Antigenic Analysis of the Filamentous
TITLE: Hemagglutinin Gene of Bordetella Pertussis
JOURNAL: Infection and Immunity
VOLUME: 58
ISSUE: 9
PAGES: 2895-2905
DATE: September 1990
US-08-465-966-16

```
alignment_scores:
  Quality: 52.50
  Ratio: 2.625
  Percent Similarity: 71.429
  Length: 28
  Gaps: 1
  Percent Identity: 46.429
```

```

alignment block:
US-09-251-133-6 x US-08-465-966-16/rev ..

Align seg 1/1 to reverse of: US-08-465-966-16 from: 1 to: 3744

      2 ThcysValLeuGlyTyrLeuHisIleValProGluPheIleGluSerGI 18
      |||||:::||||| ||::: ||:::
1925 ACCTGCATCTTGGCCCGGCGGTCGGCGGCGGCGGAGTGTGTGCGTGGC 1876
      18 nLeuLeuGlyLeuLeuSer...ProValSerLeu 28
      |||||:::||||| |||||:::|||||
1875 GCTCTTGGCCATGTGACCTGGCGGCGGCTGTGTTG 1842

seq_name: /cgn2_6/ptodata/1/1na/5A_COMB.seq:US-08-666-798-1

```

```

Sequence documentation block:
Sequence 1, Application US/08666798
Patent No. 5648238

GENERAL INFORMATION:
APPLICANT: AU-YOUNG, JANICE
APPLICANT: HAWKINS, PHILIP R.
APPLICANT: HILLMAN, JENNIFER L.
TITLE OF INVENTION: HUMAN PROTEIN KINASE C INHIBITOR HOMOLOGS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ version 1.5

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/666,798
FILING DATE: FILED HEREWITH
ATTORNEY/AGENT INFORMATION:
NAME: Luther, Barbara J.
REGISTRATION NUMBER: 33,954
REFERENCE/DOCKET NUMBER: PF-0080 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 640 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
LIBRARY: CONSENSUS
US-08-666-798-1

```

```
alignment_scores:
  Quality: 51.00      Length: 21
  Ratio: 3.188      Gaps: 0
  Percent Similarity: 76.190      Percent Identity: 47.619
```

```
alignment_block:
```

Align seg 1/1 to reverse of: US-08-666-798-1 from: 1 to: 640

```

1  ArgTyrCysValLeuGlyTyrIleuHisIleValProGluPheIleGlu 17
   |||:::|||||:::  |||:::|||||:::  ::
292  AGGAGTGTGGCTGCTGTGGGCGCATTCATCTCTCATGTGCACACT 24
17 rcInLeuengly 21

```

```
seq_name: /cgn2_6/ptodata/1/lna/5a_COMB.seq:us-08-892-692-1
242 GCAGCTCCTCGGA 230
|||||
seq_documentation_block:
; Sequence 1, Application US/08892692
; Patent No. 5773580
; GENERAL INFORMATION:
; APPLICANT: AU-YOUNG, JANICE
; APPLICANT: HAWKINS, PHILIP R.
; APPLICANT: HILLMAN, JENNIFER L.
; TITLE OF INVENTION: HUMAN PROTEIN KINASE C INHIBITOR HOMOLOG
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/892,692
; FILING DATE: 14-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/666,798
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33,954
; REFERENCE/DOCKET NUMBER: PF-0080 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 640 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; LIBRARY: CONSENSUS
; US-08-892-692-1

alignment_scores:
Quality: 51.00 Length: 21
Ratio: 3.188 Gaps: 0
Percent Similarity: 76.190 Percent Identity: 47.619

alignment_block:
US-09-251-133-6 x US-08-892-692-1/rev ..

Align seg 1/1 to reverse of: US-08-892-692-1 from: 1 to: 640
1 ArgThrcysValLeuGlyTyrLeuHisIleValProGluPheIleGlu 17
|||||
292 AGGAGTTCGCTGCGGCGCTTGGCCACTTCACTCCTCATCAGTCACACT 243
17 rGlnLeuGly 21
|||||
242 GCAGCTCCTCGGA 230

seq_name: /cgn2_6/ptodata/1/lna/5b_COMB.seq:us-09-096-071-1
seq_documentation_block:
```

```
; Sequence 1, Application US/09096071
; Patent No. 5879893
; GENERAL INFORMATION:
; APPLICANT: AU-YOUNG, JANICE
; APPLICANT: HAWKINS, PHILIP R.
; APPLICANT: HILLMAN, JENNIFER L.
; TITLE OF INVENTION: HUMAN PROTEIN KINASE C INHIBITOR HOMOLOG
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/096,071
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/666,798
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33,954
; REFERENCE/DOCKET NUMBER: PF-0080 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 640 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; LIBRARY: CONSENSUS
; US-09-096-071-1

alignment_scores:
Quality: 51.00 Length: 21
Ratio: 3.188 Gaps: 0
Percent Similarity: 76.190 Percent Identity: 47.619

alignment_block:
US-09-251-133-6 x US-09-096-071-1/rev ..

Align seg 1/1 to reverse of: US-09-096-071-1 from: 1 to: 640
1 ArgThrcysValLeuGlyTyrLeuHisIleValProGluPheIleGlu 17
|||||
292 AGGAGTTCGCTGCGGCGCTTGGCCACTTCACTCCTCATCAGTCACACT 243
17 rGlnLeuGly 21
|||||
242 GCAGCTCCTCGGA 230

seq_name: /cgn2_6/ptodata/1/lna/5b_COMB.seq:us-08-796-899-26
seq_documentation_block:
; Sequence 26, Application US/08796899
; Patent No. 6160202
; GENERAL INFORMATION:
; APPLICANT: BUSTOS, Mauricio M
; APPLICANT: CHERN, Maw-Sheng
; TITLE OF INVENTION: MODIFICATION OF SEED CROPS WITH
; TRANSCRIPTION FACTORS
```



```

: NUMBER OF SEQUENCES: 32
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Burns, Doane, Swecker & Mathis
: STREET: P.O. Box 1404
: CITY: Alexandria
: STATE: Virginia
: COUNTRY: United States
: ZIP: 22313-1404
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/796,899
: FILING DATE: 06-FEB-1997
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/319,544
: FILING DATE: 07-OCT-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Meuth, Donna M
: REGISTRATION NUMBER: 36,607
: REFERENCE/DOCKET NUMBER: 028754-005
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 836-6620
: TELEFAX: (703) 836-2021
: INFORMATION FOR SEQ ID NO: 26:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1657 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-796-899-26

alignment_scores:
  Quality: 51.00      Length: 20
  Ratio: 3.643      Gaps: 0
  Percent Similarity: 70.000      Percent Identity: 55.000

alignment_block:
  US-09-251-133-6 x US-08-796-899-26  ..

Align seg 1/1 to: US-08-796-899-26 from: 1 to: 1657

      4 ValLeuGlyTYrLeuHisIleValProGluPheIleGluSerGlnLeu 20
      ::::::::::::::||| ||| |||||
361 CTATTGGGCGACATTCACACGGTCAAGAGTTCATCCCGACCTGTCTGCT 410
20 uGlyLeuLeu 23
|||||
411 GGGACTCTCT 420

seq_name: /cgn2_6/ptodata/1/lna/6B_COWB.seq:US-09-352-159-1

seq_documentation_block:
: Sequence 1, Application US/09352159A
: Patent No. 6211434
: GENERAL INFORMATION:
: APPLICANT: DuVick, Jonathan P.
: APPLICANT: Gilliam, Jacob T.
: APPLICANT: Maddox, Joyce R.
: TITLE OF INVENTION: Amino Polyol Amine Oxidase
: TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
: FILE REFERENCE: 1134
: CURRENT APPLICATION NUMBER: US/09/352,159A
: EARLIER FILING DATE: 1999-07-12
: EARLIER APPLICATION NUMBER: 60/092,936
: EARLIER FILING DATE: 1998-07-25
: EARLIER APPLICATION NUMBER: 60/135,391
: EARLIER FILING DATE: 1999-05-21
```

```

: NUMBER OF SEQ ID NOS: 46
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 372
: TYPE: DNA
: ORGANISM: Exophiala spnifera
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (346)...(346)
: OTHER INFORMATION: n = A,T,C or G
US-09-352-159-1

alignment_scores:
  Quality: 49.00      Length: 23
  Ratio: 2.722      Gaps: 0
  Percent Similarity: 78.261      Percent Identity: 43.478

alignment_block:
  US-09-251-133-6 x US-09-352-159-1  ..

Align seg 1/1 to: US-09-352-159-1 from: 1 to: 372

      4 ValLeuGlyTYrLeuHisIleValProGluPheIleGluSerGlnLeu 20
      ::::::::::::::||| ||| |||||
14 CTCGTAGCGCTGCGCGAGTGTGTCACAGACAGACTTGTCTGCTGCTGCT 63
20 uGlyLeuLeuSerProVal 26
|||||
64 TGGACTGTGGGACCACTT 82

seq_name: /cgn2_6/ptodata/1/lna/6B_COWB.seq:US-09-352-168-1

seq_documentation_block:
: Sequence 1, Application US/09352168A
: Patent No. 6211435
: GENERAL INFORMATION:
: APPLICANT: Crasta, Oswald R.
: APPLICANT: DuVick, Jonathan P.
: APPLICANT: Folkerts, Otto
: APPLICANT: Gilliam, Jacob T.
: APPLICANT: Maddox, Joyce R.
: TITLE OF INVENTION: Amino Polyol Amine Oxidase
: TITLE OF INVENTION: Polynucleotides and Related Polypeptides and Methods of Use
: FILE REFERENCE: 0875
: CURRENT APPLICATION NUMBER: US/09/352,168A
: EARLIER FILING DATE: 1999-07-12
: EARLIER APPLICATION NUMBER: 60/092,936
: EARLIER FILING DATE: 1998-07-25
: NUMBER OF SEQ ID NOS: 33
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 372
: TYPE: DNA
: ORGANISM: Exophiala spnifera.
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (346)...(346)
: OTHER INFORMATION: n = A,T,C or G
US-09-352-168-1

alignment_scores:
  Quality: 49.00      Length: 23
  Ratio: 2.722      Gaps: 0
  Percent Similarity: 78.261      Percent Identity: 43.478

alignment_block:
  US-09-251-133-6 x US-09-352-168-1  ..

Align seg 1/1 to: US-09-352-168-1 from: 1 to: 372

      4 ValLeuGlyTYrLeuHisIleValProGluPheIleGluSerGlnLeu 20
```

```

14 CTCGTAGCTCGCGGAGTTGTCCTCCAGACGACTTTGTCTGACTGCT 63
20 UGLYLEUENSERPROVAL 26
64 TGGACTGTGGGACCACCTT 82

seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-352-159-5
seq_documentation_block:
; Sequence 5, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Gillingham, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; EARLIER FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Exophiala spintifera
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1386)
US-09-352-159-5

alignment_scores:
Quality: 49.00 Length: 23
Ratio: 2.722 Gaps: 0
Percent Similarity: 78.261 Percent Identity: 43.478

alignment_block:
US-09-251-133-6 x US-09-352-159-5/rev ..
Align seg 1/1 to reverse of: US-09-352-159-5 from: 1 to: 1389
4 ValLeuGlyTYrLeuHISileValProGluPheIleGluSerGlnLeuLe 20
:::||||| :::::|||||::: :::: ||| |||||
1125 CTCGTAGCTCGCGGAGTTGTCCTCCAGACGACTTTGTCTGACTGCT 1076
20 UGLYLEUENSERPROVAL 26
||||| :::::|||||:::
1075 TGGACTGTGGGACCACCTT 1057

seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-352-168-5
seq_documentation_block:
; Sequence 5, Application US/09352168A
; Patent No. 6211435
; GENERAL INFORMATION:
; APPLICANT: Crasta, Oswald R.
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Folkerts, Otto
; APPLICANT: Gillingham, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; FILE REFERENCE: 0875
; CURRENT APPLICATION NUMBER: US/09/352,168A
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
```

```

; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Exophiala spintifera
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1386)
US-09-352-168-5

alignment_scores:
Quality: 49.00 Length: 23
Ratio: 2.722 Gaps: 0
Percent Similarity: 78.261 Percent Identity: 43.478

alignment_block:
US-09-251-133-6 x US-09-352-168-5/rev ..
Align seg 1/1 to reverse of: US-09-352-168-5 from: 1 to: 1389
4 ValLeuGlyTYrLeuHISileValProGluPheIleGluSerGlnLeuLe 20
:::||||| :::::|||||::: :::: ||| |||||
1125 CTCGTAGCTCGCGGAGTTGTCCTCCAGACGACTTTGTCTGACTGCT 1076
20 UGLYLEUENSERPROVAL 26
||||| :::::|||||:::
1075 TGGACTGTGGGACCACCTT 1057

seq_name: /cgn2_6/ptodata/1/lna/6B_COMB.seq:US-09-352-159-10
seq_documentation_block:
; Sequence 10, Application US/09352159A
; Patent No. 6211434
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jonathan P.
; APPLICANT: Gillingham, Jacob T.
; APPLICANT: Maddox, Joyce R.
; TITLE OF INVENTION: Amino Polyol Amine Oxidase
; FILE REFERENCE: 1134
; CURRENT APPLICATION NUMBER: US/09/352,159A
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: 60/092,936
; EARLIER FILING DATE: 1998-07-25
; EARLIER APPLICATION NUMBER: 60/135,391
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 1392
; TYPE: DNA
; ORGANISM: Exophiala spintifera
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1389)
; NAME/KEY: misc_feature
; LOCATION: (1)...(3)
; OTHER INFORMATION: Extra lysine in K:trpAPO
US-09-352-159-10

alignment_scores:
Quality: 49.00 Length: 23
Ratio: 2.722 Gaps: 0
Percent Similarity: 78.261 Percent Identity: 43.478

alignment_block:
US-09-251-133-6 x US-09-352-159-10/rev ..
Align seg 1/1 to reverse of: US-09-352-159-10 from: 1 to: 1392
```


This Page Blank (uspto)


```

repeat_region 6113..6406
/note="AluYb repeat: matches 2..297 of consensus"
repeat_region 6610..6770
/note="MIR repeat: matches 42..207 of consensus"
repeat_region 6802..6852
/partial
/note="AluSp repeat: matches 1..51 of consensus"
repeat_region 6854..7046
/note="L1ME3A repeat: matches 707..908 of consensus"
repeat_region 8273..8456
/partial
/note="AluUo repeat: matches 121..302 of consensus"
tRNA join(8298..8333,8364..8404)
/note="Pro-tRNA"
/product="ad41P2A.tl"
repeat_region 8887..8926
/note="10 copies of 4 mer 85 & conserved"
repeat_region 10829..11121
/note="AluY repeat: matches 1..293 of consensus"
prim_transcript join(<11736..11849,15371..15424,25100..25192,
27966..>28026)
/note="expressed region; match: multiple ESTs: match: EST
225303 clone B7F03; match: 5' EST H53695 clone 236125"
repeat_region 12222..12399
/note="MER5B repeat: matches 178..1 of consensus"
repeat_region 12606..12787
/note="MIR repeat: matches 13..224 of consensus"
repeat_region 13061..13100
/note="20 copies of 2 mer 88 & conserved"
repeat_region 14349..14793
/partial
/note="AluX repeat: matches 37..281 of consensus"
repeat_region 16138..16433
/note="AluX repeat: matches 1..298 of consensus"
repeat_region 17588..17892
/note="AluX repeat: matches 302..1 of consensus"
repeat_region 19873..20027
/note="L1M4A repeat: matches 1047..887 of consensus"
repeat_region 20906..21174
/note="MER33 repeat: matches 41..312 of consensus"
repeat_region 21872..22372
/note="MIR repeat: matches 505..6 of consensus"
repeat_region 22890..22951
/note="MIR repeat: matches 76..142 of consensus"
repeat_region 25418..25445
/note="14 copies of 2 mer 93 & conserved"
repeat_region 28846..29145
/note="Alu repeat: matches 300..2 of consensus"
repeat_region 29874..30169
/note="AluX repeat: matches 298..1 of consensus"
repeat_region 30644..30944
/note="AluX repeat: matches 1..302 of consensus"
prim_transcript 32096..32609
/note="match: 3' EST H12785 clone 148765; Paired with EST
H12846 matching this clone"
prim_transcript complement(<32599..33072)
/note="match: 5' EST H12846 clone 148765; Paired with EST
H12785 matching this clone"
repeat_region 32761..32834
/note="2 copies of 37 mer 95 & conserved"
prim_transcript 35883..36271
/note="match: 5' EST H71901 clone 214840; Paired with EST
H71902 matching this clone"
repeat_region 37474..37501
/note="7 copies of 4 mer 93 & conserved"
prim_transcript complement(<38272..38643)
/note="match: 5' EST H71902 clone 214840; Paired with EST
H71901 matching this clone"
repeat_region 39058..39353
/note="AluY repeat: matches 1..296 of consensus"
repeat_region 39366..39667
/note="AluX repeat: matches 1..302 of consensus"
repeat_region 43904..43939

```

```

/note="18 copies of 2 mer 86 & conserved"
prim_transcript <46085..>46483
/note="match: 5' EST T86606 clone 115242"
prim_transcript <46887..>46985
/note="match: multiple ESTs"
BASE COUNT 13808 a 8845 c 8969 g 15707 t
ORIGIN

alignment_scores:
Quality: 64.00 Length: 18
Ratio: 4.000 Gaps: 0
Percent Similarity: 88.889 Percent Identity: 66.667

alignment_block:
US-09-251-133-6 x HS41P2A/rev ..

Align seg 1/1 to reverse of: HS41P2A from: 1 to: 47329

7 TyrlauhtsllevalpProgluphelegluserglneuleuglyleule 23
|||||
32037 TACCTGCATATACCTCGACATATACACTCCACTGCGCTTAAT 31988
23 user 24
||||
31987 TTCC 31984

seq_name: gb.prf:HS41P2

seq_documentation_block:
LOCUS HS41P2 108315 bp DNA PRI 12-DEC-1999
DEFINITION Human DNA sequence from clone RPI-41P2 on chromosome 22 Contains
the 3' part of the RBM9 gene for RNA binding motif protein 9 and
the 3' part of the gene for a novel protein similar to part of Apol
(apolipoprotein L) and TNF-inducible protein CG12-1. Contains
ESTs, STS and GSSs, complete sequence.
ACCESSION AL049748 281314 281357
VERSION AL049748.2 GI:6572235
KEYWORDS HTG; ApOL; apolipoprotein L; CG12-1; RBD; RBM9; RNP; RRM.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 108315)
REFERENCE
AUTHORS Whiteley,M.
TITLE Direct Submission
JOURNAL Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Dec 13, 1999 this sequence version replaced gi:4741471.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 22. constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep RPI-41P2 is from
the library RPI-1 constructed at the Roswell Park Cancer Institute
by the group of Pieter de Jong. For further details see

```

http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
IMPORTANT: This sequence is not the entire insert of clone RP1-41P2
it may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true right end of clone RP1-41P2 is at 108315 in this sequence.
The true left end of clone RP1-106120 is at 63252 in this sequence.
location/Qualifiers

FEATURES
source

1. 108315
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="RP1-41P2"
/clone_1lb="RPCR-1"
1. 160
repeat_region
/note="AluX repeat: matches 1. 159 of consensus"
1466. 1630
repeat_region
/note="AluJ repeat: matches 152. 302 of consensus"
1631. 1940
repeat_region
/note="AluX repeat: matches 1. 309 of consensus"
1941. 2086
repeat_region
/note="AluJ repeat: matches 2. 152 of consensus"
2701. 2998
repeat_region
/note="AluX repeat: matches 1. 296 of consensus"
3136. 3447
repeat_region
/note="L1M4 repeat: matches 5015. 5342 of consensus"
3450. 3831
repeat_region
/note="MER31-internal repeat: matches 2529. 3301 of consensus"
consensus"
3870. 3947
repeat_region
/note="MER31-internal repeat: matches 4484. 4561 of consensus"
4222. 4340
repeat_region
/note="MER57-internal repeat: matches 6828. 6947 of consensus"
4570. 4853
repeat_region
/note="AluSg repeat: matches 1. 310 of consensus"
4854. 4933
repeat_region
/note="20 copies 4 mer gga 86 conserved"
4875. 4930
repeat_region
/note="7 copies 8 mer gaagaa 100 conserved"
4940. 5053
repeat_region
/note="MER95 repeat: matches 42. 161 of consensus"
5930. 6082
repeat_region
/note="PRAM repeat: matches 6. 157 of consensus"
6169. 6396
repeat_region
/note="3 copies 76 mer 81 conserved"
6294. 6439
repeat_region
/note="2 copies 73 mer 84 conserved"
6477. 6624
repeat_region
/note="2 copies 74 mer 82 conserved"
6539. 6694
repeat_region
/note="6 copies 76 mer 64 conserved"
6617. 6766
repeat_region
/note="2 copies 75 mer 92 conserved"
6765. 6918
repeat_region
/note="2 copies 77 mer 86 conserved"
7038. 7367
repeat_region
/note="5 copies 66 mer 63 conserved"
7041. 7154
repeat_region
/note="6 copies 19 mer 69 conserved"
7052. 7241
repeat_region
/note="5 copies 38 mer 67 conserved"
7367. 7423
repeat_region
/note="3 copies 19 mer 84 conserved"
7429. 7653
repeat_region
/note="3 copies 75 mer 75 conserved"
7593. 7896
repeat_region
/note="4 copies 76 mer 83 conserved"
7828. 7959
repeat_region
/note="2 copies 66 mer 85 conserved"
7994. 8143

repeat_region
/note="2 copies 75 mer 88 conserved"
8072. 8223
repeat_region
/note="2 copies 76 mer 87 conserved"
8375. 8596
repeat_region
/note="3 copies 74 mer 77 conserved"
8729. 8880
repeat_region
/note="2 copies 76 mer 86 conserved"
8810. 8969
repeat_region
/note="2 copies 80 mer 83 conserved"
8898. 9122
repeat_region
/note="3 copies 75 mer 77 conserved"
9068. 9195
repeat_region
/note="2 copies 64 mer 83 conserved"
9147. 9374
repeat_region
/note="3 copies 76 mer 84 conserved"
9339. 9638
repeat_region
/note="4 copies 75 mer 72 conserved"
9644. 9943
repeat_region
/note="4 copies 75 mer 70 conserved"
10039. 10196
repeat_region
/note="2 copies 79 mer 88 conserved"
10253. 10404
repeat_region
/note="2 copies 76 mer 83 conserved"
10327. 10476
repeat_region
/note="2 copies 75 mer 85 conserved"
10483. 10632
repeat_region
/note="2 copies 75 mer 88 conserved"
10522. 10825
repeat_region
/note="4 copies 76 mer 77 conserved"
10913. 10988
repeat_region
/note="4 copies 19 mer 76 conserved"
10919. 11226
repeat_region
/note="4 copies 77 mer 65 conserved"
11289. 11513
repeat_region
/note="3 copies 75 mer 73 conserved"
11435. 11662
repeat_region
/note="3 copies 76 mer 81 conserved"
11736. 11891
repeat_region
/note="2 copies 78 mer 85 conserved"
11814. 12113
repeat_region
/note="4 copies 75 mer 72 conserved"
12309. 12460
repeat_region
/note="2 copies 76 mer 86 conserved"
12446. 12591
repeat_region
/note="2 copies 73 mer 82 conserved"
12529. 12753
repeat_region
/note="3 copies 75 mer 75 conserved"
12732. 12875
repeat_region
/note="2 copies 72 mer 84 conserved"
12825. 12976
repeat_region
/note="2 copies 76 mer 87 conserved"
12903. 13052
repeat_region
/note="2 copies 75 mer 86 conserved"
12977. 13128
repeat_region
/note="2 copies 76 mer 88 conserved"
13058. 13207
repeat_region
/note="2 copies 75 mer 90 conserved"
13129. 13432
repeat_region
/note="4 copies 76 mer 78 conserved"
13489. 13640
repeat_region
/note="8 copies 19 mer 67 conserved"
13755. 13865
repeat_region
/note="3 copies 37 mer 75 conserved"
13818. 13874
repeat_region
/note="3 copies 19 mer 84 conserved"
13953. 14102
repeat_region
/note="2 copies 75 mer 83 conserved"
14040. 14267
repeat_region
/note="3 copies 76 mer 82 conserved"
14198. 14329
repeat_region
/note="2 copies 66 mer 84 conserved"
14368. 14517
repeat_region
/note="2 copies 75 mer 88 conserved"

```

repeat_region 14446..14597
repeat_region /note="2 copies 76 mer 87 conserved"
repeat_region 14832..15001
repeat_region /note="2 copies 85 mer 83 conserved"
repeat_region 15021..15170
repeat_region /note="2 copies 75 mer 90 conserved"
repeat_region 15094..15397
repeat_region /note="4 copies 76 mer 81 conserved"
repeat_region 15373..15672
repeat_region /note="4 copies 75 mer 72 conserved"
repeat_region 15678..15977

alignment_scores:
  Quality: 64.00      Length: 18
  Ratio: 4.000        Gaps: 0
Percent Similarity: 88.889 Percent Identity: 66.667

alignment_block:
US-09-251-133-6 x HS41P2 ..

Align seg 1/1 to: HS41P2 from: 1 to: 108315

7 TyrlenuHis1leValProgiuPheIIeGluSerGlnLeuGlyLeuLe 23
|||||
6279 TACCTGCATTAACCTCCGACATATACAGCTCACTGAGCTTAAT 76328
|||||
23 user 24
76329 TTCC 76332

seq_name: gb_hcg20:AL355389

seq_documentation_block:
LOCUS AL355389 122656 bp DNA HTG 23-JAN-2001
DEFINITION Homo sapiens chromosome 1 clone RP4-740B20, *** SEQUENCING IN
PROGRES ***, 17 unordered pieces.
ACCESSION AL355389
VERSION AL355389.2 GI:9797294
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 122656)
Pavitt,R.
Direct Submission
Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Aug 12, 2000 this sequence version replaced gi:9213547.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: dj740B20
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 114021 bases at least Q40
Consensus quality: 117105 bases at least Q30
Consensus quality: 119164 bases at least Q20
Insert size: 121056; sum-of-contigs
Insert size: 139683; 7.0% error; agarose-fp
Quality coverage: 3.50x in Q20 bases; sum-of-contigs quality
coverage: 3.13x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

```

```

FEATURES
source
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2345 2444: gap of 100 bp
2445 8340: contig of 5896 bp in length
8341 8440: gap of 100 bp
8441 13714: contig of 5274 bp in length
13715 13814: gap of 100 bp
13815 17428: contig of 3614 bp in length
17429 17528: gap of 100 bp
17529 22644: contig of 5116 bp in length
22645 22744: gap of 100 bp
22745 24984: contig of 2240 bp in length
24985 25084: gap of 100 bp
25085 27553: contig of 2469 bp in length
27554 27653: gap of 100 bp
27654 43748: contig of 16095 bp in length
43749 43848: gap of 100 bp
43849 59932: contig of 16084 bp in length
59933 60032: gap of 100 bp
60033 67150: contig of 7718 bp in length
67151 67250: gap of 100 bp
67251 69496: contig of 2246 bp in length
69497 69596: gap of 100 bp
69597 77456: contig of 7860 bp in length
77457 77556: gap of 100 bp
77557 82642: contig of 5086 bp in length
82643 82742: gap of 100 bp
82743 88349: contig of 5607 bp in length
88350 88449: gap of 100 bp
88450 96215: contig of 7766 bp in length
96216 96315: gap of 100 bp
96316 106692: contig of 10377 bp in length
106693 106792: gap of 100 bp
106793 122656: contig of 15864 bp in length.
Location/Qualifiers
1..122656
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP4-740B20"
/clone_1lb="RPCI-4"
1..2344
/note="assembly_fragment:00624
fragment_chain:1"
2445..8340
/note="assembly_fragment:00921
fragment_chain:1"
8441..13714
/note="assembly_fragment:00511
fragment_chain:1"
13815..17428
/note="assembly_fragment:00485
fragment_chain:1"
17529..22644
/note="assembly_fragment:00813
fragment_chain:2"
22745..24984
/note="assembly_fragment:00507
fragment_chain:2"
25085..27553
/note="assembly_fragment:00530
fragment_chain:2"
27654..43748
/note="assembly_fragment:00004"
43849..59932
/note="assembly_fragment:00250"
60033..67150
/note="assembly_fragment:00393"
67251..69496
/note="assembly_fragment:00516"

```



```
misc_feature 69597..77456
/note="assembly_fragment:00538"
misc_feature 77557..82642
/note="assembly_fragment:00566"
misc_feature 82743..88349
/note="assembly_fragment:00737"
misc_feature 88450..96215
/note="assembly_fragment:00850"
misc_feature 96316..106692
/note="assembly_fragment:00896"
misc_feature 106793..122656
/note="assembly_fragment:00840"
/note="end:SP6"
vector_side:right"
BASE COUNT 34773 a 25217 c 27133 g 33931 t 1602 others
ORIGIN
Alignment_scores:
Quality: 64.00 Length: 24
Ratio: 3.200 Gaps: 1
Percent Similarity: 83.333 Percent Identity: 54.167
Alignment_block:
US-09-251-133-6 x AL355389/rev ..
Align seg 1/1 to reverse of: AL355389 from: 1 to: 122656
3 CysValLeuGlyTyrLeuHisIleValProGluPheIleGluSerGlnIle 19
|||||
90330 TCTATTACTATTATTTCACATCTTACCATCTCTTAGATGTAAT 90281
19 uLeu.GlyLeuLeuSerPro 25
|||||
90280 ACTCAGACTGATTCCTCCCA 90261
seq_name: gb-prf:AL139343
seq_documentation_block:
LOCUS AL139343 144818 bp DNA 09-SEP-2000
DEFINITION Human DNA sequence from clone RP5-1078M7 on chromosome 1p32.1-32.3,
complete sequence.
ACCESSION AL139343
VERSION AL139343.9 GI:10086043
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 144818)
Dunn, M.
Direct Submission
Submitted (09-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Sep 11, 2000 this sequence version replaced gi:10039534.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Bm., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
```

```
was generated from part of bacterial clone contigs of human
Chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chri
RP5-1078M7 is from the library RPCI-5 constructed at the Roswell
Park Cancer Institute by the group of Pletier de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: PCYPAC2
This sequence is the entire insert of clone RP5-1078M7. The true
left end of clone RP4-740B20 is at 74390 in this sequence. The true
right end of clone RP5-1155K23 is at 18613 in this sequence.
FEATURES
source
1..144818
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="p32.1-32.3"
/clone="RP5-1078M7"
/clone_id="RPCI-5"
complement(148..418)
/note="match: GSS: Bm:B63044"
289..741
/note="LTR40b repeat: matches 1..462 of consensus"
1044..1339
/note="AluX repeat: matches 1..298 of consensus"
2730..3043
/note="AluSg repeat: matches 1..310 of consensus"
3164..3210
/note="MIR repeat: matches 105..151 of consensus"
3573..3873
/note="AluX repeat: matches 1..306 of consensus"
3886..4199
/note="AluX repeat: matches 4..304 of consensus"
4213..4516
/note="AluSg repeat: matches 1..304 of consensus"
4734..4852
/note="AluX repeat: matches 1..304 of consensus"
4890..4980
/note="MIR repeat: matches 189..309 of consensus"
4981..5281
/note="AluY repeat: matches 1..304 of consensus"
5282..5304
/note="MIR repeat: matches 178..199 of consensus"
5305..5507
/note="MIR repeat: matches 1..230 of consensus"
5508..5561
/note="MIR repeat: matches 199..262 of consensus"
5711..6014
/note="AluSg repeat: matches 1..307 of consensus"
6413..6719
/note="AluX repeat: matches 1..307 of consensus"
7377..7591
/note="MIR repeat: matches 6..226 of consensus"
7954..8458
/note="L2 repeat: matches 2174..2708 of consensus"
8603..8906
/note="AluY repeat: matches 3..306 of consensus"
8936..9057
/note="MIR repeat: matches 71..197 of consensus"
9069..9153
/note="5 copies 17 mer 71% conserved"
9589..9780
/note="96 copies 2 mer tc 70% conserved"
9609..9780
/note="43 copies 4 mer tctc 73% conserved"
9614..9773
/note="10 copies 16 mer 75% conserved"
9619..9762
/note="12 copies 12 mer 77% conserved"
complement(9761..10266)
/note="match: GSS: Bm:AC366486"
complement(9807..10272)
/note="match: GSS: Bm:AD174667"
```

```

misc_feature      10295..10789
                   /note="match: GSS: Em:AQ473771"
repeat_region     10485..10791
                   /note="AluX repeat: matches 1. .304 of consensus"
misc_feature      11307..11423
                   /note="match: GSS: Em:B44529"
repeat_region     11526..11717
                   /note="L2 repeat: matches 2588. .2710 of consensus"
repeat_region     11768..11781
                   /note="MIR repeat: matches 31. .233 of consensus"
repeat_region     11909..11919
                   /note="L1P3 repeat: matches 5654. .5666 of consensus"
repeat_region     11984..12348
                   /note="L1P3 repeat: matches 5666. .5789 of consensus"
misc_feature      13204..13606
                   /note="L1P3 repeat: matches 5782. .6147 of consensus"
misc_feature      13204..13605
                   /note="match: GSS: Em:AQ009872"
repeat_region     13456..13467
                   /note="match: GSS: Em:AQ010332"
repeat_region     13468..13772
                   /note="MIR repeat: matches 57. .68 of consensus"
repeat_region     13773..13855
                   /note="AluSP repeat: matches 1. .310 of consensus"
repeat_region     14174..14462
                   /note="MIR repeat: matches 68. .149 of consensus"
repeat_region     14565..14873
                   /note="AluX repeat: matches 1. .290 of consensus"
repeat_region     15277..15343
                   /note="Aluub repeat: matches 1. .308 of consensus"
repeat_region     15665..15763
                   /note="MIR repeat: matches 81. .147 of consensus"
repeat_region     15783..15870
                   /note="3 copies 33 mer 82% conserved"
repeat_region     16734..16800
                   /note="2 copies 44 mer 90% conserved"
repeat_region     16852..17148
                   /note="AluJ/FLAM repeat: matches 23. .89 of consensus"
misc_feature      17200..17649
                   /note="AluSG repeat: matches 1. .295 of consensus"
repeat_region     17259..17394
                   /note="match: GSS: Em:AQ163770"
repeat_region     18947..19175
                   /note="MIR repeat: matches 107. .241 of consensus"
repeat_region     19176..19352
                   /note="L2 repeat: matches 2495. .2750 of consensus"
repeat_region     19375..19488
                   /note="MIR5B repeat: matches 1. .178 of consensus"
repeat_region     19700..20015
                   /note="L2 repeat: matches 2304. .2421 of consensus"
repeat_region     20204..20300
                   /note="AluX repeat: matches 1. .312 of consensus"
repeat_region     21153..21217
                   /note="L2 repeat: matches 2553. .2708 of consensus"
repeat_region     21648..21681
                   /note="MIR repeat: matches 112. .176 of consensus"
repeat_region     21824..22005
                   /note="17 copies 2 mer ac 82% conserved"
repeat_region     22118..22422
                   /note="MIR91A repeat: matches 6. .196 of consensus"
misc_feature      24276..24717
                   /note="AluX repeat: matches 1. .305 of consensus"
misc_feature      25148..25243
                   /note="match: GSS: Em:AQ633877"
repeat_region     25149..25216
                   /note="match: GSS: Em:AQ450746"
repeat_region     25157..25240
                   /note="match: GSS: Em:AQ334737"
repeat_region     25149..25216
                   /note="24 copies 4 mer ggga 75% conserved"
repeat_region     25157..25240
                   /note="34 copies 2 mer ag 75% conserved"

```

```

repeat_region     25162..25241
                   /note="7 copies 12 mer 78% conserved"
repeat_region     25194..25292
                   /note="5 copies 16 mer 82% conserved"
repeat_region     25196..25280
                   /note="3 copies 33 mer 79% conserved"
repeat_region     25685..25825
                   /note="5 copies 17 mer 72% conserved"
repeat_region     25826..26114
                   /note="MIR repeat: matches 8. .179 of consensus"
repeat_region     26113..26189
                   /note="AluSG repeat: matches 1. .288 of consensus"
repeat_region     26351..26472
                   /note="MIR repeat: matches 179. .245 of consensus"
repeat_region     26771..27231
                   /note="L2 repeat: matches 2577. .2710 of consensus"
repeat_region     27332..27574
                   /note="MIR repeat: matches 74. .512 of consensus"
misc_feature      complement(27558. .27964)

alignment_scores:
  Quality:      64.00      Length:      24
  Ratio:        3.200      Gaps:        1
  Percent Simlarity: 83.333      Percent Identity: 54.167

alignment_block:
  US-09-251-133-6 x AL139343 ..

Align seg 1/1 to: AL139343 from: 1 to: 144818

3 CysValLeuGlyTyrLeuHisIleValProGluPheIleGluSerGlnLe 19
|||||
96505 TGTATTACTATTATTGTCACATCTTACACATCTTACAGTGTAAATT 96554
19 uLeu.GlyLeuLeuSerPro 25
|||||
96555 ACTCAGCAGCTGATTTCCCA 96574

seq_name: gb_htg24:CNS01DRI

seq_documentation_block:
LOCUS      CNS01DRI 196023 bp DNA 25-JAN-2001
DEFINITION Homo sapiens chromosome 14 clone R-173D9, *** SEQUENCING IN
PROGRESS ***, In ordered pieces.
ACCESSION AL121594
VERSION AL121594.5 GI:12580630
KEYWORDS HTG: HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 196023)
Genoscope.
Direct Submission
Submitted (24-JAN-2001) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
On Jan 26, 2001 this sequence version replaced gi:10129660.
----- Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: Sequef@genoscope.cns.fr

IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continue. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage, etc. . even if efforts are made to eliminate these
contaminating sequences. The following BAC sequence is oriented
from the T7 to the SP6 end.

```

AUTHORS	Renner, S.S.
TITLE	Direct Subjection

```

JOURNAL      Submitted (14-FEB-1999) Biology, Univ. of Missouri-St. Louis, 8001
              Natural Bridge Rd., St. Louis, MO 63121, USA
FEATURES
  source      1..656
              /organism="Laurelia novae-zelandiae"
              /organellar="plastid:chloroplast"
              /db_xref="taxon:74879"
misc_feature  1..656
              /note="trnL-trnL intergenic spacer region"
BASE COUNT   260 a      82 c      127 g      192 t
ORIGIN
alignment_scores:
  Quality:    61.00      Length:      25
  Ratio:      3.389      Gaps:      1
  Percent Similarity: 72.000      Percent Identity: 52.000
alignment_block:
US-09-251-133-6 x AFI129032 ..
Align seg 1/1 to: AFI129032 from: 1 to: 656

1 ArgThCysValLeuGlyTyrLeuHis.....IleValProgl 13
:::|||||:::|||||:::|||||:::|||||:::|||||:::
434 AACAACATGTTATTTGGATATCTCCATCCATTTGAATTCACATTCATC 483
13 uphelleGluSerGlnLeuLeuGly 21
:::|||||:::|||||:::|||||:::|||||:::|||||
484 AATGATAGATCATCTTCTCATTTGGA 508

seq_name: gp_pll:AF129023

seq_documentation_block:
LOCUS      AFI129023      657 bp      DNA      PLN      17-MAR-1999
DEFINITION Doryphora sasafiras chloroplast trnL-trnL intergenic spacer region.
ACCESSION  AFI129023
VERSION    AFI129023.1 GI:4427015
KEYWORDS
SOURCE
ORGANISM
  Doryphora sasafiras.
  Chloroplast Doryphora sasafiras
  Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta.
  Magnoliophyta; Laurales; Monimlaceae; Doryphora.
  1 (bases 1 to 657)
REFERENCE
  Renner,S.S.
  Circumscription and phylogeny of the Laurales: evidence from
  molecular and morphological data
  Am. J. Bot. (1999) In press
  2 (bases 1 to 657)
TITLE
  Direct Submission
  Submitted (14-FEB-1999) Biology, Univ. of Missouri-St. Louis, 8001
  Natural Bridge Rd., St. Louis, MO 63121, USA
JOURNAL
FEATURES
  source      1..657
              /organism="Doryphora sasafiras"
              /organellar="plastid:chloroplast"
              /db_xref="taxon:74884"
misc_feature  1..657
              /note="trnL-trnL intergenic spacer region"
BASE COUNT   262 a      82 c      127 g      186 t
ORIGIN
alignment_scores:
  Quality:    61.00      Length:      25
  Ratio:      3.389      Gaps:      1
  Percent Similarity: 72.000      Percent Identity: 52.000
alignment_block:
US-09-251-133-6 x AFI129023 ..
Align seg 1/1 to: AFI129023 from: 1 to: 657

```

```

1  AAGGATCTAATTGGGATTCACATCATTGAATTCAGATCCATC 484
13 uphelliuserginleuLeugly 21
485 AATGATACAAATCATTTCTGATTGSA 509
seq_name: gb_D11:AF129024

seq_documentation_block:
LOCUS AF129024 657 bp DNA PLN 28-OCT-1999
DEFINITION Dryadodaphne sp. Gray 4853 chloroplast trnT-trnL intergenic spacer
region.
ACCESSION AF129024
VERSION AF129024.1 GI:4427016
KEYWORDS
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
misc_feature
BASE COUNT 261 a 83 c 125 g 188 t
ORIGIN
alignment_scores:
Quality: 61.00 Length: 25
Ratio: 3.389 Gaps: 1
Percent Similarity: 72.000 Percent Identity: 52.000
alignment_block:
US-09-251-133-6 x AF129024 ..
Align seg 1/1 to: AF129024 from: 1 to: 657

1 AAGTTCGcVallEUGLYTYrLEuHIs.....lIeValProcl 13
:::|||||:::|||||:::|||||:::|||||:::
435 AAGACAATGATATATTGGGATTCACATCATTGAATTCAGATCCATC 484
13 uphelliuserginleuLeugly 21
485 AATGATACAAATCATTTCTGATTGSA 509
seq_name: gb_pl2:AF198494

seq_documentation_block:
LOCUS AF198494 657 bp DNA PLN 29-DEC-1999
DEFINITION Doryphora aromatica chloroplast trnT-trnL intergenic spacer
sequence.
ACCESSION AF198494
VERSION AF198494.1 GI:6643923
KEYWORDS
SOURCE
Doryphora aromatica.
Chloroplast Doryphora aromatica
```

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Laurales; Monimaceae; Doryphora.

REFERENCE 1 (bases 1 to 657)
 AUTHORS Renner,S.S., Foreman,D.B. and Murray,D.
 TITLE Timing transantarctic disjunctions in the Atherospermataceae (Laurales)
 JOURNAL Syst. Biol. (2000) In press
 REFERENCE 2 (bases 1 to 657)
 AUTHORS Renner,S.S., Foreman,D.B. and Murray,D.
 TITLE Direct Submission
 JOURNAL Submitted (26-OCT-1999) Biology, University of Missouri-St. Louis, 8001 Natural Bridge Rd., St. Louis, MO 63121, USA

FEATURES
 source
 1. .657
 /organism="Doryphora aromatica"
 /organism="plastid:chloroplast"
 /db_xref="taxon:49996"

misc_feature 1. .657
 /note="trnT-trnL intergenic spacer region"

BASE COUNT 263 a 81 c 126 g 187 t
 ORIGIN

alignment_scores:
 Quality: 61.00 Length: 25
 Ratio: 3.389 Gaps: 1
 Percent Similarity: 72.000 Percent Identity: 52.000

alignment_block:
 US-09-251-133-6 x AF198494 ..

Align seg 1/1 to: AF198494 from: 1 to: 657

1 ArgThrCysValLeuGlyTyrLeuHis.....IleValProCl 13
 ::::::::::::::::::::::::::::|||:
 435 AAGACATGTATATTGGATATCTCCATCATTTGATTCAGATCCATC 484
 13 uphelliEgUserGlnLeuEngly 21
 ::::::::::::::|||:
 485 AATGATAGATCATTTCTGATTGGA 509

seq_name: gb_p11:AF129022

seq_documentation_block:
 LOCUS AF129022 668 bp DNA PLN 17-MAR-1999
 DEFINITION Daphnandra repandula chloroplast trnT-trnL intergenic spacer region.
 ACCESSION AF129022
 VERSION AF129022.1 GI:4427014
 KEYWORDS

SOURCE
 ORGANISM Daphnandra repandula.
 Chloroplast Daphnandra repandula
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Laurales; Monimaceae; Daphnandra.

REFERENCE 1 (bases 1 to 668)
 AUTHORS Renner,S.S.
 TITLE Circumscription and phylogeny of the Laurales: evidence from molecular and morphological data
 JOURNAL Am. J. Bot. (1999) In press
 REFERENCE 2 (bases 1 to 668)
 AUTHORS Renner,S.S.
 TITLE Direct Submission
 JOURNAL Submitted (14-FEB-1999) Biology, Univ. of Missouri-St. Louis, 8001 Natural Bridge Rd., St. Louis, MO 63121, USA

FEATURES
 source
 1. .668
 /organism="Daphnandra repandula"
 /organism="plastid:chloroplast"
 /db_xref="taxon:74883"

misc_feature 1. .668
 /note="trnT-trnL intergenic spacer region"

BASE COUNT 262 a 84 c 127 g 195 t
 ORIGIN

alignment_scores:
 Quality: 61.00 Length: 25
 Ratio: 3.389 Gaps: 1
 Percent Similarity: 72.000 Percent Identity: 52.000

alignment_block:
 US-09-251-133-6 x AF129022 ..

Align seg 1/1 to: AF129022 from: 1 to: 668

1 ArgThrCysValLeuGlyTyrLeuHis.....IleValProCl 13
 ::::::::::::::::::::::::::::|||:
 445 AAGACATGTATATTGGATATCTCCATCATTTGATTCAGATCCATC 494
 13 uphelliEgUserGlnLeuEngly 21
 ::::::::::::::|||:
 495 AATGATAGATCATTTCTGATTGGA 519

seq_name: gb_p5:AL135917

seq_documentation_block:
 LOCUS AL135917 104228 bp DNA PRI 02-FEB-2001
 DEFINITION Human DNA sequence from clone RP1-83M4 on chromosome 6, complete sequence.
 ACCESSION AL135917
 VERSION AL135917.15 GI:12666200
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 104228)
 AUTHORS Baguley,C.
 TITLE Direct Submission
 JOURNAL Submitted (02-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humgeny@sanger.ac.uk
 requests: clonerequests@sanger.ac.uk

On Feb 5, 2001 this sequence version replaced gi:12214260.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; SW.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
 This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr6>
 RP1-83M4 is from the library RP1-83M4-1 constructed by the group of Pletier de Jong. For further details see
<http://www.choil.org/bacpac/home.htm>
 VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone RP1-83M4 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
 The true left end of clone RP1-83M4 is at 1 in this sequence. The true left end of clone RP1-91024 is at 104129 in this sequence.

FEATURES
 source
 1. .104228
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

repeat_region /chromosome="6"
/clone="RP1-83M4"
/clone_lib="RPCL-1"
1302. .1605
/note="Alusx repeat: matches 1. .303 of consensus"
repeat_region 2223. .2533
/note="L2 repeat: matches 1866. .2190 of consensus"
repeat_region 2573. .2617
/note="MIR repeat: matches 202. .247 of consensus"
repeat_region 2581. .2634
/note="L2 repeat: matches 2693. .2747 of consensus"
repeat_region 2637. .2861
/note="MIR repeat: matches 1. .262 of consensus"
repeat_region 3966. .4024
/note="AluJ/FLAM repeat: matches 35. .77 of consensus"
repeat_region 4613. .4720
/note="27 copies 4 mer tctt 92% conserved"
repeat_region 4759. .5035
/note="AlusP repeat: matches 1. .278 of consensus"
repeat_region 5613. .5697
/note="MIR repeat: matches 65. .150 of consensus"
repeat_region 5746. .6018
/note="LIR16a repeat: matches 169. .450 of consensus"
repeat_region 6681. .6805
/note="MER5A repeat: matches 61. .185 of consensus"
repeat_region 7197. .7476
/note="2 copies 140 mer 85% conserved"
repeat_region 7544. .7704
/note="L2 repeat: matches 2355. .2530 of consensus"
repeat_region 7835. .7962
/note="L2 repeat: matches 2609. .2746 of consensus"
repeat_region 8707. .8977
/note="Alusx repeat: matches 1. .268 of consensus"
repeat_region 10041. .10339
/note="Alusx repeat: matches 1. .301 of consensus"
repeat_region 11963. .12086
/note="FLAM_C repeat: matches 1. .124 of consensus"
repeat_region 12102. .12143
/note="AluJo repeat: matches 98. .142 of consensus"
repeat_region 12144. .12456
/note="Alusx repeat: matches 1. .304 of consensus"
repeat_region 12457. .12617
/note="AluJo repeat: matches 142. .282 of consensus"
repeat_region 12648. .12786
/note="L2 repeat: matches 2596. .2745 of consensus"
repeat_region 13356. .13668
/note="AlusG repeat: matches 1. .310 of consensus"
repeat_region 13845. .14157
/note="AlusP repeat: matches 1. .313 of consensus"
repeat_region 14484. .14623
/note="MIR repeat: matches 70. .212 of consensus"
repeat_region 15188. .15497
/note="Alusx repeat: matches 1. .311 of consensus"
repeat_region 15498. .15557
/note="30 copies 2 mer cc 68% conserved"
repeat_region 15931. .16276
/note="Rigger4(Zomb1) repeat: matches 2383. .2731 of consensus"
repeat_region 16277. .16587
/note="AluDb repeat: matches 6. .312 of consensus"
repeat_region 16588. .18382
/note="Rigger4(Zomb1) repeat: matches 627. .2383 of consensus"
misc_feature 16997. .17123
/note="Sequence from overlapping clone ba3B11 (AL512354).
Assembly confirmed by restriction digest."
repeat_region 18383. .18643
/note="AluJo repeat: matches 7. .275 of consensus"
repeat_region 18644. .19263
/note="Rigger4(Zomb1) repeat: matches 1. .627 of consensus"
repeat_region 19527. .19652
/note="MIR repeat: matches 154. .258 of consensus"
repeat_region 20848. .20937

repeat_region /note="MIR repeat: matches 146. .240 of consensus"
20940. .21242
/note="LTRH repeat: matches 70. .429 of consensus"
repeat_region 22808. .23017
/note="LTRH repeat: matches 70. .284 of consensus"
repeat_region 23035. .23336
/note="AluDb repeat: matches 1. .309 of consensus"
repeat_region 23523. .23658
/note="AluDb repeat: matches 1. .139 of consensus"
repeat_region 23659. .23917
/note="AluDb repeat: matches 21. .285 of consensus"
repeat_region 24084. .24278
/note="Alusx repeat: matches 1. .195 of consensus"
repeat_region 25005. .25087
/note="MIR repeat: matches 92. .185 of consensus"
repeat_region 26139. .26440
/note="AlusG repeat: matches 1. .299 of consensus"
repeat_region 26889. .27041
/note="MIR repeat: matches 78. .252 of consensus"
repeat_region 27623. .28017
/note="THE1B repeat: matches 1. .364 of consensus"
repeat_region 29032. .29383
/note="THE1A repeat: matches 1. .354 of consensus"
repeat_region 29985. .30289
/note="Alusx repeat: matches 1. .307 of consensus"
repeat_region 30313. .30457
/note="MER5A repeat: matches 2. .189 of consensus"
repeat_region 32411. .33100
/note="L1PA7 repeat: matches 5426. .6134 of consensus"
repeat_region 33301. .33442
/note="MIR repeat: matches 71. .232 of consensus"
repeat_region 34707. .34752
/note="23 copies 2 mer tt 76% conserved"
repeat_region 34819. .34990
/note="L1PA5 repeat: matches 5972. .6143 of consensus"
repeat_region 34992. .35031
/note="10 copies 4 mer aaat 85% conserved"
misc_feature 35362. .36232
/note="CPG Island"
/evidence=not_experimental
repeat_region 37387. .37697
/note="AluY repeat: matches 1. .311 of consensus"
repeat_region 37710. .37993
/note="AluJo repeat: matches 1. .286 of consensus"
repeat_region 38849. .39089
/note="MIR repeat: matches 2. .262 of consensus"
repeat_region 40658. .40945
/note="Alusx repeat: matches 1. .288 of consensus"
repeat_region 41221. .41256
/note="18 copies 2 mer ac 100% conserved"
repeat_region 41267. .41347
/note="L2 repeat: matches 1806. .1889 of consensus"
repeat_region 41323. .41464
/note="L2 repeat: matches 2007. .2143 of consensus"
repeat_region 41487. .41557
/note="L1MB2 repeat: matches 6099. .6167 of consensus"
repeat_region 41558. .41965
/note="MST1 repeat: matches 1. .426 of consensus"
repeat_region 41966. .42346
/note="L1MB2 repeat: matches 5705. .6099 of consensus"
repeat_region 42352. .42498
/note="L2 repeat: matches 2172. .2318 of consensus"
repeat_region 42500. .42591
/note="23 copies 4 mer tctt 75% conserved"
repeat_region 42501. .42588
/note="44 copies 2 mer tt 73% conserved"
repeat_region 42592. .42760
/note="FRAM repeat: matches -1. .176 of consensus"
repeat_region 43997. .44315
/note="Alusx repeat: matches 1. .309 of consensus"
repeat_region 45082. .45396
/note="AlusG repeat: matches 1. .313 of consensus"
repeat_region 45830. .46141

This sequence was finished as follows unless otherwise noted:

Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,

Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, N., Nguyen, N., Nickerson, E., Nwokewo, S., Ogund, A., Okunola, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Perez, J., Peters, L., Peters, L., Pickens, R., Primus, E., P. L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojurokan, I., Rolfe, M., Ruiz, S., Severy, G., Scherer, S., Scott, G., Shen, H., Shooshari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, R., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Thomas, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleczky, R., Wooten, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zortilla, S., Nelson, D., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 156768)
Worley, K.C.

Direct Submission
Submitted (23-MAY-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 9, 2000 this sequence version replaced gi:9966656.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: HBNW
Center clone name: RP11-592B21
----- Summary Statistics -----
Sequencing vector: M13; L08821
Chemistry: Dye-primer Bodipy: 74% of reads
Assembly program: Phrap; version 0.90329
Consensus quality: 148444 bases at least Q40
Consensus quality: 153372 bases at least Q30
Consensus quality: 155446 bases at least Q20
Estimated insert size: 153985; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 4.8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see <http://www.hgsc.bcm.tmc.edu/docs/Genbankdraftdata.html>).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 53909: contig of 53909 bp in length
* 53910 54009: gap of unknown length
* 54010 81912: contig of 27903 bp in length
* 81913 82012: gap of unknown length
* 82013 98592: contig of 16580 bp in length
* 98593 98692: gap of unknown length
* 98693 113882: contig of 14690 bp in length
* 113883 113482: gap of unknown length
* 113483 125818: contig of 12336 bp in length
* 125819 125918: gap of unknown length
* 125919 136097: contig of 10179 bp in length
* 136098 136197: gap of unknown length
* 136198 143818: contig of 7621 bp in length
* 143819 143918: gap of unknown length
* 143919 149744: contig of 5826 bp in length
* 149745 149844: gap of unknown length
* 149845 152407: contig of 2553 bp in length
* 152408 152507: gap of unknown length
* 152507 155627: contig of 3120 bp in length
* 155628 155727: gap of unknown length

FEATURES
source 155728 156768: contig of 1041 bp in length.
1. 156768
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-592B21"

BASE COUNT 50062 a 27554 c 27078 g 51071 t 1003 others
ORIGIN

alignment_scores
quality: 61.00 length: 24
ratio: 3.211 gaps: 0
Percent Similarity: 79.167 Percent Identity: 54.167

alignment_block:
us-09-251-133-6 x AC069263/rev ..

Align seg 1/1 to reverse of: AC069263 from: 1 to: 156768

4 ValLeuGlyTyrLeuHisLeValProGluPheIleGluSerGlnLeuLe 20
152009 GTGCTGTGCTCTTGCCTTAATTCAGAACTTTGGAATCTCAGACAGC 151960
151959 TGGATTCCTTGACGCTCTGAGT 151938

20 uGlyLeuLeuSerProValSer 27
seq_name: gb_hlg21:AL356986

seq_documentation_block:
LOCUS AL356986 161082 bp DNA HTG 15-APR-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-395N6, *** SEQUENCING IN
PROGRESS ***, 33 unordered pieces.
ACCESSION AL356986
VERSION AL356986.6 GI:12331052
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 161082)
Plumb, B.
Submitted (14-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Jan 22, 2001 this sequence version replaced gi:9797641.

----- Genome Center -----
Center: Sanger Centre
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquerry@sanger.ac.uk
----- Project Information -----
Center project name: BA395N6
----- Summary Statistics -----
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 144239 bases at least Q40
Consensus quality: 150252 bases at least Q30
Consensus quality: 153770 bases at least Q20
Insert size: 157882; sum-of-contigs
Insert size: 206767; 2.8% error; agarose-fp
Quality coverage: 2.85x in Q20 bases; sum-of-contigs Quality
coverage: 2.53x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 5857: contig of 5857 bp in length
5858 5957: gap of 100 bp
5958 8262: contig of 2305 bp in length
8263 8362: gap of 100 bp
8363 10908: contig of 2546 bp in length
10909 11008: gap of 100 bp
11009 14014: contig of 3006 bp in length
14015 14114: gap of 100 bp
14115 16904: contig of 2790 bp in length
16905 17004: gap of 100 bp
17005 20121: contig of 3117 bp in length
20122 20221: gap of 100 bp
20222 24722: contig of 4501 bp in length
24723 24822: gap of 100 bp
24823 30221: contig of 5399 bp in length
30222 30321: gap of 100 bp
30322 33337: contig of 3016 bp in length
33338 33437: gap of 100 bp
33438 38178: contig of 4741 bp in length
38179 38278: gap of 100 bp
38279 42633: contig of 4355 bp in length
42634 42733: gap of 100 bp
42734 61499: contig of 18766 bp in length
61500 61599: gap of 100 bp
61600 75243: contig of 13650 bp in length
75250 75349: gap of 100 bp
75350 78034: contig of 2685 bp in length
78035 78134: gap of 100 bp
78135 82107: contig of 3973 bp in length
82108 82207: gap of 100 bp
82208 85543: contig of 3342 bp in length
85549: gap of 100 bp
85550 87894: contig of 2245 bp in length
87895 87994: gap of 100 bp
87995 90215: contig of 2221 bp in length
90216 90315: gap of 100 bp
90316 93124: contig of 2809 bp in length
93125 93224: gap of 100 bp
93225 96890: contig of 3666 bp in length
96891 96990: gap of 100 bp
96991 100946: contig of 3956 bp in length
100947 101046: gap of 100 bp
101047 104990: contig of 3944 bp in length
104991 105090: gap of 100 bp
105091 115155: contig of 10065 bp in length
115156 115255: gap of 100 bp
115256 118113: contig of 2858 bp in length
118114 118213: gap of 100 bp
118214 124762: contig of 6549 bp in length
124763 124862: gap of 100 bp
124863 129730: contig of 4868 bp in length
129731 129830: gap of 100 bp
129831 133308: contig of 3478 bp in length
133309 133408: gap of 100 bp
133409 142799: contig of 9391 bp in length
142800 142899: gap of 100 bp
142900 149102: contig of 6203 bp in length
149103 149202: gap of 100 bp
149203 151847: contig of 2645 bp in length
151848 151947: gap of 100 bp
151948 156179: contig of 4232 bp in length
156180 156279: gap of 100 bp
156280 158289: contig of 2010 bp in length
158290 158389: gap of 100 bp
158390 161082: contig of 2693 bp in length.

FEATURES

source
1. .161082
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"

misc_feature
1. 5857
/note="assembly_fragment:00141
fragment_chain:1"
misc_feature
5958. 8262
/note="assembly_fragment:00959
fragment_chain:1"
misc_feature
8363. 10908
/note="assembly_fragment:01368
fragment_chain:1"
misc_feature
11009. 14014
/note="assembly_fragment:00059
fragment_chain:2"
misc_feature
14115. 16904
/note="assembly_fragment:00570
fragment_chain:2"
misc_feature
17005. 20121
/note="assembly_fragment:00728
fragment_chain:3"
misc_feature
20222. 24722
/note="assembly_fragment:01417
fragment_chain:3"
misc_feature
24823. 30221
/note="assembly_fragment:00836
fragment_chain:4"
misc_feature
30322. 33337
/note="assembly_fragment:01228
fragment_chain:4"
misc_feature
33438. 38178
/note="assembly_fragment:01243
fragment_chain:5"
misc_feature
38279. 42633
/note="assembly_fragment:01509
fragment_chain:5"
misc_feature
42734. 61499
/note="assembly_fragment:01251
fragment_chain:6"
misc_feature
61600. 75249
/note="assembly_fragment:01359
fragment_chain:6"
misc_feature
75350. 78034
/note="assembly_fragment:01505
fragment_chain:7"
misc_feature
78135. 82107
/note="assembly_fragment:00026
fragment_chain:7"
misc_feature
82208. 85549
/note="assembly_fragment:00001"
fragment_chain:7"
misc_feature
85650. 87894
/note="assembly_fragment:00042"
fragment_chain:7"
misc_feature
87995. 90215
/note="assembly_fragment:00072"
fragment_chain:7"
misc_feature
90316. 93124
/note="assembly_fragment:00137"
fragment_chain:7"
misc_feature
93225. 96890
/note="assembly_fragment:00198"
fragment_chain:7"
misc_feature
96991. 100946
/note="assembly_fragment:00350"
fragment_chain:7"
misc_feature
101047. 104990
/note="assembly_fragment:00515"
fragment_chain:7"
misc_feature
105091. 115155
/note="assembly_fragment:00643"
fragment_chain:7"
misc_feature
115256. 118113
/note="assembly_fragment:00701"
fragment_chain:7"
misc_feature
118214. 124762
/note="assembly_fragment:00710"
fragment_chain:7"
misc_feature
124863. 129730
/note="assembly_fragment:00711"
fragment_chain:7"
misc_feature
129831. 133308
/note="assembly_fragment:00771"
fragment_chain:7"
misc_feature
133409. 142799
/note="assembly_fragment:00841"
fragment_chain:7"

misc_feature 142900..149102 /note="assembly_fragment:01110.0"
misc_feature 149203..151847 /note="assembly_fragment:01172"
misc_feature 151948..156179 /note="assembly_fragment:01431"
misc_feature 156280..158289 /note="assembly_fragment:01493"
misc_feature 158390..161082 /note="assembly_fragment:01334
clone_end:SP6
vector_side:right"

BASE COUNT 43396 a 35726 c 35099 g 43643 t 3218 others

alignment_scores:
Quality: 61.00 Length: 20
Ratio: 3.588 Gaps: 0
Percent Similarity: 85.000 Percent Identity: 50.000

alignment_block:
US-09-251-133-6 x AL356986 ..

Align seg 1/1 to: AL356986 from: 1 to: 161082

9 HistyleValProGIupheIlleGIuSerGIuLeuGIyLeuSerPr 25
||||:||||| |||:||||:||||: |||:||||:||||:
43394 CATGTATCCCGACGATTCTGTGAGGCTGACATGAGGATGCTTGAGCC 43443
25 ovalSerLeu 28
| |||||
43444 CAGAGTTTG 43453

FT	/tag- x	/bound_moleity= "Primer C5"
FT	/note= "Amplification of amplicon 9-30"	
FT	misc.feature	10990..11442
FT	/tag- y	/note= "Amplicon 99-14387"
FT	primer_bind	10990..11008
FT	/tag- z	/bound_moleity= "Primer B6"
FT	/note= "Amplification of amplicon 99-14387"	
FT	primer_bind	11423..11442
FT	/tag- aa	/note= "Amplification of amplicon 99-14387"
FT	misc.feature	12472..12966
FT	/tag- ab	/note= "Amplicon 99-14389"
FT	primer_bind	12472..12491
FT	/tag- ac	/bound_moleity= "Primer B7"
FT	/note= "Amplification of amplicon 99-14389"	
FT	primer_bind	12946..12966
FT	/tag- ad	/bound_moleity= "Primer C7"
FT	/note= "Amplification of amplicon 99-14389"	
FT	misc.feature	15073..15320
FT	/tag- ae	/note= "Amplicon 9-12"
FT	primer_bind	15073..15092
FT	/tag- af	/bound_moleity= "Primer B8"
FT	/note= "Amplification of amplicon 9-12"	
FT	primer_bind	15503..15520
FT	/tag- ag	/bound_moleity= "Primer C8"
FT	/note= "Amplification of amplicon 9-12"	
FT	misc.feature	15311..15551
FT	/tag- ah	/note= "Amplicon 9-13"
FT	primer_bind	15311..15150
FT	/tag- ai	/bound_moleity= "Primer B9"
FT	/note= "Amplification of amplicon 9-13"	
FT	primer_bind	15532..15551
FT	/tag- aj	/bound_moleity= "Primer C9"
FT	/note= "Amplification of amplicon 9-13"	
FT	misc.feature	15759..16211
FT	/tag- ak	/note= "Amplicon 99-14405"
FT	primer_bind	15759..15776
FT	/tag- al	/bound_moleity= "Primer B10"
FT	/note= "Amplification of amplicon 99-14405"	
FT	primer_bind	16191..16211
FT	/tag- am	/note= "Amplification of amplicon 99-14405"
FT	misc.feature	16233..16652
FT	/tag- an	/note= "Amplicon 9-14"
FT	primer_bind	16233..16251
FT	/tag- ao	/bound_moleity= "Primer B11"
FT	/note= "Amplification of amplicon 9-14"	
FT	primer_bind	16653..16652
FT	/tag- ap	/bound_moleity= "Primer C11"
FT	/note= "Amplification of amplicon 9-14"	
FT	misc.feature	16604..17005
FT	/tag- aq	/note= "Amplicon 9-15"
FT	primer_bind	16604..16621

```

FT      /tag= ar
FT      /bound_moiety= "Primer B12"
FT      /note= "Amplification of amplicon 9-15"
FT      primer_bind
FT      17006..17025
FT      /tag= as
FT      /bound_moiety= "Primer C12"
FT      /note= "Amplification of amplicon 9-15"
FT      misc_feature
FT      16982..17402
FT      /tag= at
FT      /note= "Amplicon 9-16"
FT      primer_bind
FT      16982..17001
FT      /tag= au
FT      /bound_moiety= "Primer B13"
FT      /note= "Amplification of amplicon 9-16"
FT      primer_bind
FT      17384..17402
FT      /tag= av
FT      /bound_moiety= "Primer C13"
FT      /note= "Amplification of amplicon 9-16"
FT      misc_feature
FT      17216..17517
FT      /tag= aw
FT      /note= "Amplicon 9-17"
FT      primer_bind
FT      17216..17233
FT      /tag= ax
FT      /bound_moiety= "Primer B14"
FT      /note= "Amplification of amplicon 9-17"
FT      primer_bind
FT      17498..17517
FT      /tag= ay
FT      /bound_moiety= "Primer C14"
FT      /note= "Amplification of amplicon 9-17"
FT      misc_feature
FT      17300..17503

alignment_scores:
      Quality: 56.00      Length: 29
      Ratio: 2.800      Gaps: 1
Percent Similarity: 68.966      Percent Identity: 41.379

alignment_block:
US-09-251-133-6 x AAD00552/rev ..

Align seg 1/1 to reverse of: AAD00552 from: 1 to: 20966

      1 ArghrHcysValIleuGIYrLeu.....HisIleValProG1 13
      ::::::::::: ::::::::::: |||||:::|
      6958 AAGGTTCCTCCCTACAGCCACCTTAATAAAGCAGCACATATATATCCAGC 6909
      13 uPhelIeGIuSerGIuLeuGIuLeuSerPro 25
      ||| |||::: ::::: ||||| |||||
      6908 CTTTGGGAGGCTGAGTGGGAGAGACTGCTGAGCCCA 6872

seq_name: /SID58/gcgcdata/geneseq/geneseqn/NA1999.DAT:AAZ233902

seq_documentation_block:
ID      AAZ233902 standard; DNA: 49999 BP.
XX
XX      AAZ233902;
XX
XX
XX      25-JAN-2000 (first entry)
XX
XX      Human LOBO homologue genomic DNA fragment 4.
DE
XX      LOBO; long bones; bone development; bone extension; skull; osteopathic;
XX      diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
XX      spondyloepiphyseal dysplasia; achondroplasia; human; ds.
XX
XX      Homo sapiens.
XX
XX      WO950284-A2.
XX
XX      07-OCT-1999.
XX
XX      26-MAR-1999.      99WO-EP02055.
XX
XX      27-MAR-1998.      98DE-1013799.
XX

```

XX (ROSE/) ROSENTHAL A.
PA
XX Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
PI
XX MPI, 1999-601320/51.
DR
XX
PT Nucleic acids encoding proteins which influence bone development,
PT useful for treating and studying bone disorders -
XX
PS
XX Example 3; Page 300-328; 391pp; German.

XX This invention describes novel nucleic acids (I; designated LOBO (long
CC bones)) encoding proteins influencing bone development in mammals. The
CC proteins of the invention reduce and/or inactivate bone extension (i.e.
CC development), with exception of the skull and have osteopathic activity.
CC The nucleic acid molecules, proteins and antibodies can be used in
CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC and nucleic acid molecules, etc. are useful for production of transgenic
CC animals, especially a transgenic mouse for the study of diseases
CC associated with bone development, e.g. spondyloepiphyseal dysplasia and
CC achondroplasia. This sequence encodes a human LOBO protein described
XX in the method of the invention.
XX

XX Sequence 49999 BP; 12459 A; 12933 C; 12356 G; 12251 T; 0 other;
SQ

[illegible]

alignment_block:
US-09-251-133-6 x AAZ23902/rev

Align seg 1/1 to reverse of: AAZ23902 from: 1 to: 499999

```

2  ThcysVal.....LeuGlyTyrLeuHisIleVal)ProGluPh 14
   ::::: ::::: ::::: :::::
41460 TCCTGATTTTTTTTATAGAGATGGGGTTTTGCCATGTTGCCACAGCTTA 414111

```

14 eilegluSerGlnLeuLeuGlyLeuLeuSerPro 25
:::||||| ||||| |||||
41410 TCCTGAACTCCAGCTACTGGGCTCAAGCAGTCCA 41377

seq_name: /SIDS8/gcgydata/geneseq/geneseqn/NA1998.DAT:AAV40370

```
seq_documentation_block:
ID AAV40370 standard; DNA; 3744 BP.
```

AC	AAV40370;
XX	
DT	07-DEC-1998 (first entry)

DE Filamentous haemagglutinin fragment 7 gene.

KW Filamentous haemagglutinin; FHA; endothelial cell;
KW Integrin receptor; blood-brain barrier; antiinflammatory;
KW Inflammation; meningitis; therapy; whooping cough; vaccine; ss

OS Bordetella pertussis.

PN US5792457-A.

PD 11-AUG-1998.

PF 03-MAY-1991; 91US-06955613.

PR 30-NOV-1994; 94US-0348353.

PR 04-MAY-1992; 92WO-US03725.

PR U6-JUN-1995; 95US-0465929.

XX
PA (UVRQ) UNIV ROCKEFELLER

PI Masure HR, Tuomanen E;

DR WPI; 1998-456074/39

XX
X

Page 10

PT filamentous haemagglutinin RGD regions

PS Disclosure; Flg 10A-L; 64pp; English.

CC This DN sequence codes for 'fragment 7' (see AAM659594) of Bordetella
CC pertussis filamentous haemagglutinin (FHA). Fragment 7 includes
CC the RGD tripeptide that corresponds to amino acid residues 1057-1099
CC of FHA, and a carbohydrate recognition site corresponding to amino
CC acid residues 1141-1279 of FHA. Truncated FHAs (see AAM659595)
CC which delete the RGD region can be produced genetically and are
CC useful as vaccines against whooping cough. The invention also
CC provides peptides and antibodies which inhibit the reaction between
CC the RGD tripeptide of FHA and the integrin receptors of endothelial
CC cells and their utility as therapeutic agents, as well as a method
CC of increasing the permeability of the blood-brain barrier using an
CC antibody to the FHA RGD region.

SQ Sequence 3744 BP; 741 A; 1157 C; 1328 G; 518 T; 0 other;

[illegible]

alignment_block:

US-09-251-133-6 x AAV40370/rev .

Align seg 1/1 to reverse of: AAV40370 from: 1 to: 3744

```

2 ThrcysValLeuGlyTyrLeuHisIleValProGluPheIleGluSerG1 18
  |||||:::||||| ||::: ::||| ||:::
1925 ACCTGCATCTTGCCCGCGCTGCGCGCGGCCCGGAGTCTCAGCGTGCC 1876

```

```

18 nleuLeuGlyleuLeuSer..ProValSerleu 28
      |||||:::  ||| |||||
1875 GCTCTTGCCCATGACCTCGGCCCCGGTCTGTTG 1842

```

seq_name: /SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ30127

```
seq_documentation_block:
```

ID	AAZ30127	standard; DNA; 3744 BP.
XX		
AC	AAZ30127;	

DT 26-JAN-2000 (first entry)

DE Nucleic acid encoding fragment 7 of filamentous hemagglutinin.

KW Filamentous hemagglutinin; FHA; Factor X; Leukocyte;

KW non-fimbrial surface associated protein; *Bordetella pertussis*; C3b1;

KW infection; meningitis; septic arthritis; endophthalmitis;

XX

XX

FT	CDS	1.3744
----	-----	--------

```

ET
/notes= "partial sequence"

```

PN US5968512-A.

KW Infection; meningitis; septic arthritis; endophthalmitis; ss.
 OS Bordetella pertussis.
 XX
 XX US5932217-A.
 PN
 PD 03-AUG-1999.
 XX
 PF 30-NOV-1994; 94US-0348353.
 XX
 XX 30-NOV-1994; 94US-0348353.
 PR 03-MAY-1991; 91US-0695613.
 PR 04-MAY-1992; 92RO-US03725.
 PR 23-MAY-1994; 94US-0247572.
 XX
 XX (UYRQ) UNIV ROCKEFELLER.
 XX
 PI Measure HR, Tuomanen E;
 DR WPI; 1999-443571/37.
 DR P-PsDB; AAY23960.
 XX
 XX Peptides inhibiting the adhesion between leukocytes and endothelial
 PT cells, useful for treating inflammation
 PS
 PS Disclosure; Fig 10A-L; 82pp; English.
 XX
 XX The specification describes peptides which inhibit adhesion between
 CC leukocytes and endothelial cells, and peptides which inhibit adhesion
 CC between bacteria (Bordetella pertussis) and ciliated respiratory
 CC epithelial cells. The specification also describes an immunogenic
 CC composition comprising a polypeptide portion of Bordetella pertussis
 CC filamentous haemagglutinin (FHA) containing no RGD region or containing
 CC an amino acid sequence altered in the RGD region, where the polypeptide
 CC portion elicits antibodies which do not cross-react with cerebral
 CC endothelial cells. The peptides and methods are useful for reducing
 CC inflammation during the course of antibiotic therapy of infectious
 CC diseases such as meningitis, septic arthritis, and endophthalmitis.
 CC The present sequence encodes fragment 7 of the FHA protein of
 CC B. pertussis.
 XX
 XX Sequence 3744 BP; 741 A; 1147 C; 1338 G; 518 T; 0 other;
 SQ
 Alignment_scores:
 Quality: 52.50 Length: 28
 Ratio: 2.625 Gaps: 1
 Percent Similarity: 71.429 Percent Identity: 46.429
 Alignment_block:
 US-09-251-133-6 x AAX86180/rev ..
 Align seg 1/1 to reverse of: AAX86180 from: 1 to: 3744
 2 ThrCysValLeuGIYTrLeuHisIleValProGluPheIleGluSerGI 18
 |||||:::||||| ||::: ::||| ||:::
 1925 ACCTGCACATCTTCGCCGCGTTCGCGCGCCGCCGGAAGTCGTACGCTGGC 1876
 18 nLeuLeuGlyLeuLeuSer...ProValSerLeu 28
 |||||:::||||| ||||| ||||| ||||| |||||
 1875 GCTCTTGGCCATGACCTCGCGCCGCCGCTCTCGTTG 1842
 seq_name: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ57049
 seq_documentation_block:
 ID AAZ57049 standard; DNA: 3744 BP.
 XX AAZ57049;
 AC
 XX 30-MAY-2000 (first entry)
 DT
 XX FHA fragment 7 polypeptide encoding gene.
 DE
 XX

```
KM FHA; blood-brain barrier; filamentous hemagglutinin; endothelial cell;
KM complement C3b1; factor X; integrin receptor CR3; leukocyte migration;
KM bacterial adhesion; brain cancer; acquired immune deficiency syndrome;
KM Parkinson's disease; Alzheimer's disease; antibacterial; anti-epileptic;
KM anti-inflammatory; anticancer; antiviral; antineurodegeneration; ds.
XX
XX Bordetella pertussis.
OS
XX US6015560-A.
XX
XX 18-JAN-2000.
PD
XX
XX 06-JUN-1995; 95US-0465966.
PF
XX
XX 30-NOV-1994; 94US-0348353.
PR
XX 04-MAY-1992; 92MO-US03725.
PR 03-MAY-1991; 91US-0695613.
PR 23-MAY-1994; 94US-0247572.
XX
XX (UYRQ ) UNIV ROCKEFELLER.
PA
XX Measure HR, Tuomanen E;
PI
XX WPI: 2000-181133/16.
DR
XX P-PSDB; AAY67527.
XX
XX Compositions containing antibody to filamentous hemagglutinin, used
PT e.g. to increase permeability of the blood-brain barrier and to inhibit
PT inflammation or bacterial adhesion -
XX
XX Disclosure; Fig 10A-L; 82pp; English.
PS
XX The invention provides a novel pharmaceutical composition for increasing
CC the permeability of the blood-brain barrier to a molecule (1). The
CC composition comprises (1) and an antibody to FHA (filamentous
CC hemagglutinin) which binds to endothelial cells in brain blood vessels,
CC increasing permeability. FHA contains polypeptide regions with binding
CC properties similar to those of complement C3b1, factor X and integrin
CC receptor CR3, and some anti-FHA antibodies are competitive inhibitors
CC of these materials, i.e. they reduce leukocyte migration or bacterial
CC adhesion. The compositions are used to improve delivery of (1) to the
CC brain, e.g. where (1) is used to treat brain cancer, acquired immune
CC deficiency syndrome, epilepsy, Parkinson's or Alzheimer's diseases or
CC other neurological diseases. Other antibodies directed against particular
CC regions of FHA are used to treat inflammation (caused by microbial
CC infection or auto-immune disease), also to prevent adhesion of Bordetella
CC pertussis to respiratory endothelial cells. The present sequence
CC represents a FHA fragment 7 encoding gene.
XX
XX Sequence 3744 BP; 741 A; 1147 C; 1338 G; 518 T; 0 other;
SQ
XX
XX alignment_scores:
XX Quality: 52.50 Length: 28
XX Ratio: 2.625 Gaps: 1
XX Percent Similarity: 71.429 Percent Identity: 46.429
XX
XX alignment_block:
XX US-09-251-133-6 x AA257049/rev ..
XX
XX Align seg 1/1 to reverse of: AA257049 from: 1 to: 3744
XX
XX 2 ThcCysValLeuGlyTyrLeuHisIleValProGluPheIleGluSerG1 18
XX |||||:||||| |||||:||||| |||||:||||| |||||:|||||
XX 1925 ACCGTCATCTTGGCCGCGTGGCCGCGGCCCGGAAAGTCTCAAGCGTGC 1876
XX
XX 18 nLeuLeuGlyLeuLeuSer...ProValSerLeu 28
XX |||||:||||| |||||:||||| |||||:||||| |||||:|||||
XX 1875 GCTCTGGCATGACCTCGGCGCCGCGTCTGTTG 1842
XX
XX seq_name: /SID88/gcgdata/geneseq/geneseqn/AA2000.DAT:AAA10263
XX
XX seq_documentation_block:
```

```
ID AAA10263 standard; DNA; 11883 BP.
XX
XX AAA10263:
AC
XX
XX 03-JUL-2000 (first entry)
XX
XX Bordetella pertussis filamentous haemagglutinin gene, fhaB.
DE
XX
XX Filamentous haemagglutinin; FHA; fhaB gene; surface-associated;
KM cellular adhesion; antigenic; immunogenic; whooping cough; pertussis;
KM diagnosis; prophylaxis; vaccine; ds.
XX
XX Bordetella pertussis.
OS
XX
XX Key Location/Qualifiers
FH 253..11043
FT CDS /"tag" 8
FT /product= "B. pertussis filamentous haemagglutinin"
XX
XX US6036960-A.
XX
XX 14-MAR-2000.
PD
XX
XX 01-SEP-1994; 94US-0299941.
XX
XX 13-NOV-1989; 89US-0436297.
XX 10-AUG-1992; 92US-0928619.
XX 27-OCT-1988; 88US-0263648.
XX
XX (RELM/) REIMAN D A.
XX (DOME/) DOMENIGHINI M.
XX (RAPP/) RAPPUOLI R.
XX (FALK/) FALKOW S.
XX
XX Falow S, Domenighini M, Reiman DA, Rappuoli R;
PI WPI: 2000-269872/23.
DR
XX P-PSDB; AAY87407.
XX
XX New isolated nucleic acid encoding the filamentous hemagglutinin of
PT Bordetella pertussis, useful for diagnosis, treatment and prevention of
PT whooping cough -
XX
XX Claim 3; Columns 11-28; 17pp; English.
XX
XX This sequence represents the filamentous haemagglutinin gene, fhaB, from
CC Bordetella pertussis. Bordetella pertussis is a small Gram negative
CC bacillus found only in humans and which causes whooping cough (pertussis)
CC in children. Filamentous haemagglutinin (FHA) is a 368 kD surface-
CC associated protein which is thought to be one of the most important
CC factors in mediating the adhesion of the bacterium to a eukaryotic
CC (human) cell. FHA also stimulates an immune response in humans following
CC disease, and has been shown to act as an immunoprotective antigen in an
CC animal model. The FHA gene, or its expression products, are used for
CC prevention or treatment of pertussis, particularly in vaccines. It may
CC also be used to increase the amount of FHA in live or dead Bordetella
CC pertussis or other organisms. Fragments of the gene may be used as
CC diagnostic hybridisation probes or as antisense modulators to reduce the
CC infectivity of Bordetella pertussis. FHA, or fragments thereof, may be
CC used in immunoassays to detect anti-FHA antibodies, and to raise
CC antibodies for the diagnosis, prevention of treatment of whooping cough.
XX
XX Sequence 11883 BP; 2196 A; 3648 C; 4282 G; 1754 T; 3 other;
SQ
XX
XX alignment_scores:
XX Quality: 52.50 Length: 28
XX Ratio: 2.625 Gaps: 1
XX Percent Similarity: 71.429 Percent Identity: 46.429
XX
XX alignment_block:
XX US-09-251-133-6 x AAA10263/rev ..
XX
```



```
XX 08-JUL-1999.
PD
XX
XX 22-DEC-1998; 98WO-US27598.
XX
XX 31-DEC-1997; 97US-0002485.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Baughn MR, Corley NC, Guegler KJ, Hillman JL, Lal P;
PI Sather SK, Shah P;
XX
XX WPI; 1999-430242/36.
DR
XX P-PSDB; AA921841.
XX
XX Human signal-peptide containing protein coding sequences used to
PT treat cancer and immune responses
XX
XX Claim 9; Page 89; 99pp; English.
XX
XX The invention provides human signal-peptide containing proteins (SIGP)
CC (AA921841-855) and polynucleotides (AA92076-90) encoding the proteins.
CC A host cell containing a vector comprising SIGP DNA can be used to
CC produce the SIGP protein. The SIGP protein can be used, in conjunction
CC with a pharmaceutical carrier to treat or prevent a cancer. An antagonist
CC of the SIGP protein can be used to treat or prevent a cancer or an
CC immune response. The cancers that can be treated or prevented include
CC sarcomas, adenocarcinomas, leukemias, lymphomas, melanomas,
CC teratocarcinomas, myelomas and cancers of the adrenal gland, bladder,
CC bone, bone marrow, brain, breast, cervix, gall bladder, ganglia,
CC gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary,
CC pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen,
CC testis, thymus, thyroid, and uterus. The immune responses that can be
CC treated or prevented include, AIDS, Addison's disease, adult respiratory
CC distress syndrome, allergies, anemia, asthma, atherosclerosis,
CC bronchitis, cholecystitis, Crohn's disease, ulcerative colitis, atopic
CC dermatitis, dermatomyositis, diabetes mellitus, emphysema, atrophic
CC gastritis, glomerulonephritis, Grave's disease, gout, hyperosiphilia,
CC irritable bowel syndrome, lupus erythematosus, multiple sclerosis,
CC myasthenia gravis, myocardial or pericardial inflammation,
CC osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid
CC arthritis, scleroderma, Sjogren's syndrome, and autoimmune thyroiditis,
CC complications of cancer, infections, and trauma.
XX
XX Sequence 846 BP; 273 A; 186 C; 210 G; 177 T; 0 other;
SQ
alignment_scores:
Quality: 50.50 Length: 26
Ratio: 2.525 Gaps: 1
Percent Similarity: 76.923 Percent Identity: 38.462
alignment_block:
US-09-251-133-6 x AA92076/rev ..
Align seg 1/1 to reverse of: AA92076 from: 1 to: 846
1 ArgThrCysValIeu.....GLTYrLeuHsIleValProGluPh 14
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
484 AAGCCCTGCATCCTGTTCTTGATTAATTTCCTGCTGTCGCGGATTT 435
14 eileGluSerGlnLeuLeuGlyLeuLeu 23
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
434 CATCAAAACCAAGCCCTGTCCTC 407
seq_name: /stds/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA81477
seq_documentation_block:
ID AAA81477 standard; DNA; 26778 BP.
XX
XX AAA81477;
XX
XX 08-DEC-2000 (first entry)
XX
```

```
XX
DE N. meningitidis partial DNA sequence gnm_25 SEQ ID NO:25.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
XX antigen; vaccine; diagnosis; infection; antibacterial; identification;
XX Meningococcus B; MenB; ds.
XX
XX Neisseria meningitidis.
XX
XX WO200022430-A2.
XX
XX 20-APR-2000.
XX
XX 08-OCT-1999; 99WO-US23573.
XX
XX 09-OCT-1998; 98US-0103794.
XX 30-APR-1999; 99US-0132068.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarcelll M, Scarlato V;
PI Rappoli R, Pizzi M;
XX
XX WPI; 2000-318079/27.
XX
XX Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX
XX Claim 7; Page 524-531; 1760pp; English.
XX
XX The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AA81453 to AA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences; AA81260 to AA81303 and AA825620 to AA825663 represent
CC Neisseria DNA sequences and their corresponding proteins; AA81254 to
CC AA81259 and AA81304 to AA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AA81322 to
CC AA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against Meningococcus B; against all serotypes;
CC and/or against all pathogenic Neisseriae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC Multivalent vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface
CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
XX other more variable regions.
XX
XX Sequence 26778 BP; 6602 A; 7008 C; 6846 G; 6318 T; 4 other;
SQ
alignment_scores:
Quality: 50.50 Length: 26
Ratio: 3.156 Gaps: 1
Percent Similarity: 61.538 Percent Identity: 46.154
alignment_block:
US-09-251-133-6 x AAA81477/rev ..
Align seg 1/1 to reverse of: AAA81477 from: 1 to: 26778
3 CysValLeuGlyTYrLeuHsIleValProGluPh..... 14
||| ||||||||| |||:||||| |||:|
```

21978 TGCTTCTTGTTATTACACGTAAGAGATACGAAATACATCA 21929
15 .IleGluSerGlnLeuGlyLeu 23
::: |||::||| ||| |||
21928 CACAAATTACAAATCTGTTGCTG 21901

This Page Blank (usptc)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2001, 13:24:06 ; Search time 12.41 Seconds
(without alignments)
46.457 Million cell updates/sec

Title: US-09-251-133-6

Perfect score: 141
1 RTCVLGYLHIVEFIESQLGLSPVSL 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCITUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51	36.2	552	4 US-08-796-899-27	Sequence 27, Appl
2	49	34.8	570	1 US-08-403-866-6	Sequence 6, Appl
3	47	33.3	308	1 US-08-457-245-8	Sequence 8, Appl
4	45	31.9	822	2 US-08-939-002A-16	Sequence 16, Appl
5	44	31.2	139	4 US-09-382-155-20	Sequence 20, Appl
6	44	31.2	139	4 US-09-074-044A-20	Sequence 20, Appl
7	44	31.2	237	4 US-08-818-514-3	Sequence 3, Appl
8	44	31.2	237	4 US-09-115-934A-3	Sequence 3, Appl
9	44	31.2	372	1 US-08-403-634-2	Sequence 30, Appl
10	44	31.2	372	1 US-08-403-634-30	Sequence 2, Appl
11	44	31.2	372	4 US-08-913-441B-2	Sequence 2, Appl
12	44	31.2	372	4 US-08-913-441B-30	Sequence 30, Appl
13	44	31.2	2595	4 US-09-036-887A-2	Sequence 2, Appl
14	43	30.5	326	6 5171684-7	Patent No. 5171684
15	43	30.5	535	2 US-08-933-750C-20	Sequence 20, Appl
16	43	30.5	535	4 US-09-234-613-20	Sequence 20, Appl
17	42	29.8	662	1 US-07-841-651-4	Sequence 4, Appl
18	40	28.4	429	1 US-08-339-152A-33	Sequence 33, Appl
19	40	28.4	429	5 PCT-US95-11808-4	Sequence 4, Appl
20	40	28.4	429	5 PCT-US95-11808-4	Sequence 4, Appl
21	40	28.4	451	1 US-08-191-337-3	Sequence 3, Appl
22	40	28.4	705	2 US-08-456-647B-4	Sequence 4, Appl
23	40	28.4	705	2 US-08-237-201A-4	Sequence 4, Appl
24	40	28.4	974	3 US-08-938-291A-4	Sequence 4, Appl
25	40	28.4	1090	3 US-08-307-896-3	Sequence 3, Appl
26	40	28.4	1090	3 US-08-726-214-4	Sequence 4, Appl
27	40	28.4	1090	5 PCT-US95-11808-3	Sequence 3, Appl

28	39.5	28.0	358	1 US-08-239-431A-4	Sequence 4, Appl
29	39.5	28.0	358	2 US-08-463-081B-6	Sequence 6, Appl
30	39.5	28.0	358	2 US-08-461-379A-6	Sequence 6, Appl
31	39.5	28.0	358	2 US-08-462-390B-6	Sequence 6, Appl
32	39.5	28.0	358	3 US-08-463-074B-6	Sequence 6, Appl
33	39.5	28.0	358	3 US-08-465-585C-6	Sequence 6, Appl
34	39.5	28.0	358	3 US-08-652-446-6	Sequence 6, Appl
35	39.5	28.0	553	3 US-09-083-351-2	Sequence 2, Appl
36	39.5	28.0	553	4 US-09-083-352-2	Sequence 2, Appl
37	39.5	27.7	346	3 US-08-988-111-2	Sequence 2, Appl
38	39.5	27.7	346	4 US-09-387-922-2	Sequence 2, Appl
39	39.5	27.7	447	4 US-09-378-255-2	Sequence 2, Appl
40	39.5	27.7	475	4 US-09-251-372-2	Sequence 2, Appl
41	39.5	27.7	541	2 US-08-484-438-6	Sequence 6, Appl
42	39.5	27.7	574	4 US-09-552-351-2	Sequence 2, Appl
43	39.5	27.7	880	2 US-08-916-917-12	Sequence 12, Appl
44	39.5	27.7	880	3 US-09-225-170-12	Sequence 12, Appl
45	39.5	27.7	880	4 US-09-141-212-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1

US-08-796-899-27

Sequence 27, Application US/08796899

Patent No. 6160202

GENERAL INFORMATION:

APPLICANT: BUSTOS, Mauricio M

INVENTOR: CHERN, Maw-Sheng

TITLE OF INVENTION: MODIFICATION OF SEED CROPS WITH

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESSES:

ADDRESS: Burns, Doane, Swecker & Mathis

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/796,899

FILING DATE: 06-FEB-1997

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/319,544

FILING DATE: 07-OCT-1994

ATTORNEY/AGENT INFORMATION:

NAME: Neuh, Donna M

REGISTRATION NUMBER: 36,607

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 27:

SEQUENCE CHARACTERISTICS:

LENGTH: 552 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-796-899-27

Query Match 36.2%, Score 51, DB 4, Length 552;
Best Local Similarity 55.0%, Pred. No. 3.2;
Matches 11, Conservative 3, Mismatches 6, Gaps 0;

```
QY      4 VLGYLHVPEFIESQLLGLL 23
          :||:| | | | | | | |
DB     121 LGHINTVKEFHPHLLGLL 140
```

RESULT 2
US-08-403-

```

US-08-403-866--6
Sequence 6, Application US/08403866
Patent No. 5643779
GENERAL INFORMATION:
APPLICANT: Ehrlich, Stanislaw
APPLICANT: Godon, Jean-Jacques
APPLICANT: Renault, Pierre
TITLE OF INVENTION: Nucleic acid coding for an alpha-acetolactate
TITLE OF INVENTION: synthase from Lactococcus and its applications
NUMBER OF SEQUENCES: 16
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403,866
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 20747/30
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1487
TELEFAX: (716) 263-1600
TELEX: 978450 (WUT)
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 570 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Lactococcus lactis subsp. lactis
INDIVIDUAL ISOLATE: ILVD
US-08-403-866--6

```

Query Match	34.88;	Score 49;	DB 1;	Length 570;
Best Local Similarity	60.08;	Pred. No. 7;		
Matches	9;	Conservative	3;	Mismatches 3;
				Indels 0;
				Gaps 0;

```

QY      9 HIVPEFIESQLGLL 23
         ||||| :| |::
Db     495 HIVPEAVEGGLIGLV 509

```

RESULT 3
US-08-457-245-8

GENERAL INFORMATION: CLONING AND EXPRESSION OF DNA INVOLVED
APPLICANT: YUAN, YING IN THE BIOSYNTHESIS OF CYCLOPROPANATED MICOLIC ACIDS IN
TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS
TITLE OF INVENTION: 21
NUMBER OF SEQUENCES: 21

```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: Floppy disk
3  COMPUTER: IBM PC compatible
4  OPERATING SYSTEM: PC-DOS/MS-DOS
5  SOFTWARE: PatentIn Release #1.0, Version #1.30
6  CURRENT APPLICATION DATA:
7  APPLICATION NUMBER: US/08/457,245
8  FILING DATE:
9  CLASSIFICATION: 435
10 ATTORNEY/AGENT INFORMATION:
11 NAME: Chabbers, Guy W.
12 REGISTRATION NUMBER: 30,617
13 REFERENCE/DOCKET NUMBER: 15280-216000
14 TELECOMMUNICATION INFORMATION:
15 TELEPHONE: (415) 543-9600
16 TELEFAX: (415) 543-5043
17 INFORMATION FOR SRD ID NO: 8:
18 SEQUENCE CHARACTERISTICS:
19 LENGTH: 308 amino acids
20 TYPE: amino acid
21 STRANDEDNESS: not relevant
22 TOPOLOGY: not relevant
23 MOLECULE TYPE: protein
24 ORIGINAL SOURCE:
25 ORGANISM: Mycobacterium leprae
26 US-08-457-245-8

```

Query Match	33.38;	Score 47;	DB 1;	Length 308;
Best Local Similarity	52.28;	Pred. No. 7.3;		
Matches 12;	Conservative 5;	Mismatches 4;	Indels 2;	Gaps 1

```
QY      8 LH--IVPEFIESQLGLLSPVSL 28
        || |||: ||: |||: ||: ||
Db      187 LHTIIVPDAKETKELGLTTPMSL 209
```

RESULT 4

US-08-939-002A-16
Sequence 16, Application US/08939002A
Patent No. 5845529
GENERAL INFORMATION:
APPLICANT: HAYASHI, KIYOSHI
APPLICANT: LIU, AIMIN
APPLICANT: LI, HEBIAO
APPLICANT: HARAGUCHI, KAZUNOMO
APPLICANT: KITAMURA, YOSHIKI
TITLE OF INVENTION: CELLOBIOSE PHOSPHORYLASE GENE, VECTOR
TITLE OF INVENTION: AND TRANSFORMANT CONTAINING SAID GENE
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/939,002A
FILING DATE: 26-SEP-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 221193/1997
FILING DATE: 04-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618


```

? REFERENCE/DOCKET NUMBER: 8361-001-0
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 703-413-3000
? TELEFAX: 703-413-2220
? INFORMATION FOR SEQ. ID NO: 16:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 822 amino acids
? TYPE: amino acid
? TOPOLOGY: 1linear
? MOLECULE TYPE: protein
? US-08-939-002A-16

```

Query Match	31.9%	Score 45:	DB 2:	Length 822:
Best Local Similarity	28.6%	Pred. No.	46:	
Matches	6:	Conservative	9:	Mismatches
			6:	Indels
			0:	Gaps
			0:	

```
OY      4  VLGLHVPEFIESQLGLLS 24
          :|::|::||      :::  |
Db     373  LGGVHLPERARERIIDIAS 393
```

```

RESULT      5
US-09-382-155-20
: Sequence 20, Application US/09382155B
: Patent No. 6160095
: GENERAL INFORMATION:
: APPLICANT: CHAUDHARY, PREET M
: APPLICANT: HOOD, LEROY
: TITLE OF INVENTION: PROTEINS CAPABLE OF REGULATING NF-KB, JNK AND APOPTOSIS
: TITLE OF INVENTION: PATHWAYS AND METHODS OF USING THE SAME
: FILE REFERENCE: Chaudhary
: CURRENT APPLICATION NUMBER: US/09/382,155B
: CURRENT FILING DATE: 1999-08-24
: EARLIER APPLICATION NUMBER: 09/074,044
: EARLIER FILING DATE: 1998-05-07
: NUMBER OF SEQ ID NOS: 40
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 20
: LENGTH: 139
: TYPE: PRT
: ORGANISM: Molluscum contagiosum virus
: US-09-382-155-20

```

Query Match	31.28;	Score 44;	DB 4;	Length 139;
Best Local	Similarity 47.18;	Pred. No. 9;		
Matches	8;	Conservative 3;	Mismatches 6;	Indels 0;
			Gaps	0;

```
QY      8 LHVPEFIESQLGLLS 24
        ||:| |:| | | :|
Db      73 LHLDPFLEERHLGTM 89
```

RESULT 6
US-09-074-044A-20
Sequence 20. Application US/09074044A
Patent No. 6207458
GENERAL INFORMATION:
APPLICANT: CHAUDHARY, PREET M
APPLICANT: HOOD, LEROY
TITLE OF INVENTION: PROTEINS CAPABLE OF REGULATING NK-KB, JNK AND
TITLE OF INVENTION: APOPTOSIS PATHWAYS AND METHODS OF USING THE SAME
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
STREET: 2405 GRAND BLVD., SUITE 400
CITY: KANSAS CITY
STATE: MISSOURI
COUNTRY: USA
ZIP: 64108
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/074,044A
FILING DATE: 09/07/93

Query Match	31.2%	Score	44	DB	4	Length	139
Best Local Similarity	47.1%	Pred. No.	9				
Matches	8	Conservative	3	Mismatches	6	Indels	0
						Gaps	0

```
QY      8 LHVPEFIESQLGLLS 24
          ||: ||:| | | :|
Db      73 LHLDPFLERHLA GTMS 89
```

RESULT 7
 US-08-818-514-3
 Sequence 3, Application US/08818514
 Patent No. 5837838
 GENERAL INFORMATION:
 APPLICANT: Reed, John C.
 APPLICANT: Xu, Qunli
 TITLE OF INVENTION: BAX Inhibitor Proteins
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Campbell & Flores
 STREET: 4370 La Jolla Village Drive, Suite 7000
 City: San Diego
 STATE: California
 COUNTRY: United States
 ZIP: 92122
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/818,514
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-LJ 2446
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 237 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-818-514-3

Query Match 31.2%; Score 44; DB 2; Length 237;
Best Local Similarity 40.9%; Pred. No. 16;
Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 7 YLHIVEPFIESQLGLSPVSL 28
|:|:| |:|:| |:|:|
DB 46 YVHMVTHFIQAGLSALGSLL 67

RESULT 8

US-09-115-934A-3
; Sequence 3, Application US/09115934A
; Patent No. 6130317
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Xu, Qunli
; TITLE OF INVENTION: BAX Inhibitor Proteins
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,934A
; FILING DATE:
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/818,514
; FILING DATE: 14-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 3209
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-09-115-934A-3

Query Match 31.2%; Score 44; DB 4; Length 237;
Best Local Similarity 40.9%; Pred. No. 16;
Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 7 YLHIVEPFIESQLGLSPVSL 28
|:|:| |:|:| |:|:|
DB 46 YVHMVTHFIQAGLSALGSLL 67

RESULT 9

US-08-403-634-2
; Sequence 2, Application US/08403634
; Patent No. 5674748
; GENERAL INFORMATION:
; APPLICANT: Giordano, Antonio
; TITLE OF INVENTION: NOVEL HUMAN CYCLIN-DEPENDENT
; KINASE-LIKE PROTEINS AND METHODS
; OF USING THE SAME
; NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESSEE: No. 5674748ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA

QY 6 GYHIVEPFIESQLGLSPV 26
|:|:| |:|:| |:|:|
DB 97 GSIVLDFCEHDLGLSNV 117

US-08-403-634-2
; Sequence 30, Application US/08403634
; Patent No. 5674748
; GENERAL INFORMATION:
; APPLICANT: Giordano, Antonio
; TITLE OF INVENTION: NOVEL HUMAN CYCLIN-DEPENDENT
; KINASE-LIKE PROTEINS AND METHODS
; OF USING THE SAME
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESSEE: No. 5674748ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,634
; FILING DATE:
; CLASSIFICATION: 424
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/208,575
; FILING DATE: 08-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJD-1482
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 372 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-403-634-2

Query Match 31.2%; Score 44; DB 1; Length 372;
Best Local Similarity 47.6%; Pred. No. 27;
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 6 GYHIVEPFIESQLGLSPV 26
|:|:| |:|:| |:|:|
DB 97 GSIVLDFCEHDLGLSNV 117

US-08-403-634-30
; Sequence 30, Application US/08403634
; Patent No. 5674748
; GENERAL INFORMATION:
; APPLICANT: Giordano, Antonio
; TITLE OF INVENTION: NOVEL HUMAN CYCLIN-DEPENDENT
; KINASE-LIKE PROTEINS AND METHODS
; OF USING THE SAME
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESSEE: No. 5674748ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,634
; FILING DATE:
; CLASSIFICATION: 424
; PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US 08/208,575
FILING DATE: 08-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: 700-1482
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-403-634-30

Query Match 31.2%; Score 44; DB 1; Length 372;
Best Local Similarity 47.6%; Pred. No. 27;
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 6 GYLHIVEPFIESQLGLSPV 26
|:::|:| | | | | | | | | |
Db 97 GSILVDFCEHDLAGLLSNV 117

RESULT 11
US-08-913-441B-2
Sequence 2, Application US/08913441B
Patent No. 6162612
GENERAL INFORMATION:
APPLICANT: Giordano, Antonio
TITLE OF INVENTION: No. 6162612el Human Cyclin-Dependent Kinase-Like Proteins and
TITLE OF INVENTION: Methods of Using The Same
FILE REFERENCE: 8321-76 C11
CURRENT APPLICATION NUMBER: US/08/913,441B
CURRENT FILING DATE: 1997-12-02
PRIOR APPLICATION NUMBER: 08/403,634
PRIOR FILING DATE: 1995-03-14
PRIOR APPLICATION NUMBER: PCT/US96/03557
PRIOR FILING DATE: 1996-03-14
NUMBER OF SEQ ID NOS: 31
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 2
LENGTH: 372
TYPE: PRT
ORGANISM: Homo sapiens
US-08-913-441B-2

Query Match 31.2%; Score 44; DB 4; Length 372;
Best Local Similarity 47.6%; Pred. No. 27;
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 6 GYLHIVEPFIESQLGLSPV 26
|:::|:| | | | | | | | | |
Db 97 GSILVDFCEHDLAGLLSNV 117

RESULT 12
US-08-913-441B-30
Sequence 30, Application US/08913441B
Patent No. 6162612
GENERAL INFORMATION:
APPLICANT: Giordano, Antonio
TITLE OF INVENTION: No. 6162612el Human Cyclin-Dependent Kinase-Like Proteins and
TITLE OF INVENTION: Methods of Using The Same
FILE REFERENCE: 8321-76 C11
CURRENT APPLICATION NUMBER: US/08/913,441B
CURRENT FILING DATE: 1997-12-02
PRIOR APPLICATION NUMBER: 08/403,634
PRIOR FILING DATE: 1995-03-14

PRIOR APPLICATION NUMBER: PCT/US96/03557
PRIOR FILING DATE: 1996-03-14
NUMBER OF SEQ ID NOS: 31
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 30
LENGTH: 372
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:
US-08-913-441B-30

Query Match 31.2%; Score 44; DB 4; Length 372;
Best Local Similarity 47.6%; Pred. No. 27;
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 6 GYLHIVEPFIESQLGLSPV 26
|:::~|:~ | | | | | | | | | |
Db 97 GSILVDFCEHDLAGLLSNV 117

RESULT 13
US-09-036-987A-2
Sequence 2, Application US/09036987A
Patent No. 6143526
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Patli J.
APPLICANT: Turner, Jan R.
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
TITLE OF INVENTION: Production
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dow Agrosciences LLC Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036,987A
FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R.
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2595 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-036-987A-2

Query Match 31.2%; Score 44; DB 4; Length 2595;
Best Local Similarity 55.6%; Pred. No. 2,4e+02;

Matches 10; Conservative 3; Mismatches 3; Indels 2; Gaps 1;
QY 9 HIVEPIESQLGLSPV 26
| | | : | | | | : | |
Db 724 HT--EVVRDELGLAPV 739

RESULT 14
5171684-7
; Patent No. 5171684
; APPLICANT: YEN, KWANG-MU; BLATT, LAWRENCE M.; KARL, MICHAEL R.
; TITLE OF INVENTION: BIOCONVERSIONS CATALYZED BY THE TOULUENE
; MONOOXYGENASE OF PSEUDOMONAS MENDOCINA KR-1
; NUMBER OF SEQUENCES: 41
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/590,374
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 177,631
; FILING DATE: 05-APR-1988
; SEQ ID NO: 7:
; LENGTH: 326
5171684-7

Query Match 30.5%; Score 43; DB 6; Length 326;
Best Local Similarity 58.8%; Pred. No. 34;
Matches 10; Conservative 4; Mismatches 1; Indels 2; Gaps 2;

QY 6 GYLH-IVPEIESQLG 21
| | | | : | | | : | |
Db 275 GYVHEIIPPEYMEG-LLG 290

RESULT 15
US-08-933-750C-20
; Sequence 20, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750C
; FILING DATE: September 23, 1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 535 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAITTT03
; CLONE: 864683
US-08-933-750C-20

Query Match 30.5%; Score 43; DB 2; Length 535;
Best Local Similarity 45.5%; Pred. No. 59;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
QY 3 CVLGYLHIVE 13
| | | | : | | | : | |
Db 340 CIVGHQHWIPE 350

Search completed: August 7, 2001, 13:29:00
Job time: 294 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2001, 13:28:41 ; Search time 23.02 Seconds
(without alignments)
160.927 Million cell updates/sec

Title: US-09-251-133-6

Perfect score: 141
1 RTCVLGYLHIVEPIESQLGLLSPVSL 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SPREMBL_16:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rhodent:*
13: sp_unclassified:*
14: sp_vertebrate:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	36.9	239	8 Q9MS96	Q9MS96 galdieria s
2	52	36.9	688	5 Q9VXB7	Q9VXB7 galdieria s
3	50	35.5	812	2 052504	052504 thermotoga
4	50	35.5	813	2 087964	087964 thermotoga
5	50	35.5	813	2 Q9X263	Q9X263 thermotoga
6	49	34.8	275	2 006111	006111 clostridium
7	48.5	34.4	469	5 Q9VEP9	Q9VEP9 drosophila
8	48	34.0	99	6 Q9VZU3	Q9VZU3 sus scrofa
9	48	34.0	512	5 Q9VZU3	Q9VZU3 sus scrofa
10	48	34.0	626	5 Q9U3X4	Q9U3X4 dictyostell
11	48	34.0	702	1 Q9YF08	Q9YF08 aeropyrum p
12	47	33.3	146	10 Q9LRT3	Q9LRT3 arabidopsis
13	47	33.3	308	2 Q49807	Q49807 mycobacteri
14	47	33.3	308	3 Q04919	Q04919 saccharomyc
15	47	33.3	532	1 027545	027545 methanobact
16	46.5	33.0	415	6 018856	018856 caenolestes
17	46.5	33.0	550	10 Q9LIM4	Q9LIM4 eucalyptus
18	46.5	33.0	554	5 Q93174	Q93174 caenorhabdi
19	46.5	33.0	681	2 Q9LIX1	Q9LIX1 streptomyc

20	46.5	33.0	913	5 Q97272	Q97272 plasmodium
21	46	32.6	169	10 Q9S5V1	Q9S5V1 arabidopsis
22	46	32.6	179	2 083125	083125 treponema p
23	46	32.6	662	10 Q9M9T8	Q9M9T8 arabidopsis
24	46	32.6	964	11 Q9QX51	Q9QX51 mus musculu
25	46	32.6	4684	4 Q15149	Q15149 homo sapien
26	45.5	32.3	395	6 Q18857	Q18857 phascogale
27	45.5	32.3	425	6 Q18859	Q18859 dromiclops
28	45.5	32.3	323	6 Q18860	Q18860 vomatius ur
29	45.5	32.3	643	10 Q9LR59	Q9LR59 arabidopsis
30	45.5	32.3	871	10 Q9YCK9	Q9YCK9 aeropyrum p
31	45.5	32.3	2048	10 Q9SIF1	Q9SIF1 arabidopsis
32	45	31.9	288	2 P96312	P96312 burkholderi
33	45	31.9	310	14 Q98240	Q98240 molinuscum c
34	45	31.9	388	2 Q67472	Q67472 aquilex aeo
35	45	31.9	405	5 Q9U3D1	Q9U3D1 caenorhabdi
36	45	31.9	409	5 P91783	P91783 polyorchis
37	45	31.9	428	2 Q9EY64	Q9EY64 vibrio chol
38	45	31.9	432	10 Q04473	Q04473 arabidopsis
39	45	31.9	477	2 Q9KRT5	Q9KRT5 pseudomonas
40	45	31.9	480	2 Q9F8X3	Q9F8X3 pseudomonas
41	45	31.9	558	3 P87161	P87161 aspergillus
42	45	31.9	811	2 Q59316	Q59316 clostridium
43	45	31.9	822	2 Q66264	Q66264 celivibrio
44	45	31.9	1013	14 Q9Q078	Q9Q078 chimpanzee
45	45	31.9	1319	10 Q9SLE9	Q9SLE9 arabidopsis

ALIGNMENTS

RESULT 1					
ID Q9MS96	PRELIMINARY:	PRT:	239 AA.		
AC Q9MS96;					
DT 01-OCT-2000 (TREMBLrel. 15, Created)					
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)					
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)					
DE HYPOTHETICAL 26.8 KDA PROTEIN.					
OS Galdieria sulphuraria.					
OC Chloroplast.					
OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;					
OC Galdieria.					
OX NCBI_Taxid=130081;					
RN [1]					
RP SEQUENCE FROM N.A.					
RC STRAIN=UTEX 2393;					
RL Whitney S.M., Andrews J.;					
DR Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.					
DR EMBL: AF233069; AAF81686.1; -					
DR InterPro: IPR001687; -					
DR InterPro: IPR003439; -					
DR InterPro: IPR003593; -					
DR Pfam: PF00005; ABC_tran; 1.					
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.					
DR SMART: SM00382; AAA; 1.					
KW Hypothetical protein; Chloroplast.					
SO SEQUENCE 239 AA; 26836 MW; 58A2868B54A85FD CRC64;					
Query Match	36.9%;	Score 52;	DB 8;	Length 239;	
Best Local Similarity	44.0%;	Pred. No. 7.1;			
Matches 11;	Conservative 44;	Mismatches 7;	Indels 2;	Gaps 1;	
OY 1 RTCVLGYLHIVEPIESQLGLLSP 25					
DB 14 RKCLIDNLHMT--FSQYEILGLGP 36					
RESULT 2					
ID Q9VXB7	PRELIMINARY:	PRT:	688 AA.		
AC Q9VXB7;					

```
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-JUN-2000 (TReMBLrel. 13, Last sequence update)
DE CG4724 PROTEIN.
GN CG4724.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN
[1]
SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Abmayyan A., An H.-T., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burks K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Deyes A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz S., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,
RA Jajalil M., Kalush F., Karpen G.H., Ke Z., Kensington J.A., Ketchem K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mlshina N.V., Modarres C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleib J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003503; AAF48660.1; -.
DR FlyBase: Fgn0030785; CG4724.
DR InterPro: IPR002554; -.
DR Pfam: P01603; B56; 1.
SQ
SEQUENCE 688 AA; 74530 MW; F5C3E6FDD0180F74 CRC64;
```

```
Query Match 36.9%; Score 52; DB 5; Length 688;
Best Local Similarity 42.9%; Pred. No. 22;
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;
```

```
QY 3 CVGLHIVPEFIESOLLGLL 23
DB 164 CITAFLNQPSTFKYKGLL 184

RESULT 3
ID 052504 PRELIMINARY; PRT; 812 AA.
AC 052504;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
```

```
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE CELLOBIOSE PHOSPHORYLASE (EC 2.4.1.20).
GN CBPA.
OS Thermotoga neapolitana.
OC Bacteria; Thermotogales; Thermotoga.
NCBI_TaxID=2337;
RN
[1]
SEQUENCE FROM N.A.
RA Yernool D.A., Bok J.D., McCarthy J.K., Sullivan R.F., Eveleigh D.E.;
RT "Cloning and characterization of gluco-oligosaccharide catabolic
RT pathway: beta-glucan glucosylidase and cellobiose phosphorylase, in
RT the marine hyperthermophile Thermotoga neapolitana.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF039487; AAB95491.2; -.
DR InterPro: IPR000169; -.
DR PROSITE: PS00639; THIO_L_PROTEASE_HIS; UNKNOWN_1.
KW transferase; Glycosyltransferase.
SQ
SEQUENCE 812 AA; 93515 MW; DED324725B2CCDE CRC64;
```

```
Query Match 35.5%; Score 50; DB 2; Length 812;
Best Local Similarity 38.1%; Pred. No. 52;
Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 4 VLGLHIVPEFIESOLLGLL 24
DB 366 ILGFVHMIPKARQRIIDLAS 386

RESULT 4
ID 087964 PRELIMINARY; PRT; 813 AA.
AC 087964;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE CELLOBIOSE-PHOSPHORYLASE (EC 2.4.1.20).
GN CBPA.
OS Thermotoga neapolitana.
OC Bacteria; Thermotogales; Thermotoga.
NCBI_TaxID=2337;
RN
[1]
SEQUENCE FROM N.A.
RC STRAIN=Z2706-MC24;
RA Zverlov V., Bronnenmeier K.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN
[2]
SEQUENCE FROM N.A.
RC STRAIN=Z2706-MC24;
RX MEDLINE=93356813; PubMed=8352795;
RA Dakhova O., Kureplina N., Zverlov V., Svetlichny V.,
RA Velikodvorskiya G.;
RT "Cloning and expression in Escherichia coli of Thermotoga neapolitana
RT genes coding for enzymes of carbohydrate substrate degradation.";
RL Biochem. Biophys. Res. Commun. 194:1359-1364(1993).
DR EMBL: Z99777; CAB16926.1; -.
DR InterPro: IPR000169; -.
DR PROSITE: PS00639; THIO_L_PROTEASE_HIS; UNKNOWN_1.
KW transferase; Glycosyltransferase.
SQ
SEQUENCE 813 AA; 93677 MW; 04DAFF39E61AEA88 CRC64;
```

```
Query Match 35.5%; Score 50; DB 2; Length 813;
Best Local Similarity 38.1%; Pred. No. 52;
Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 4 VLGLHIVPEFIESOLLGLL 24
DB 367 ILGFVHMIPKARQRIIDLAS 387

RESULT 5
```

09X2G3 ID 09X2G3 PRELIMINARY; PRT; 813 AA.
 AC 09X2G3:
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
 DE CELLOBIOSE-PHOSPHORYLASE.
 GN TM1848.
 OS Thermotoga maritima.
 CC Bacteria; Thermotogales; Thermotoga.
 OK NCBI_TaxID=2336;
 RP [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MSB8 / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
 Halt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.,
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 RT genome sequence of Thermotoga maritima."
 RL Nature 399:323-329(1999).
 DR EMBL; AE001822; AAD36910.1; -
 DR TIGR; TM1848; -
 DR InterPro: IPR000169; -
 DR PROSITE; PS00639; THIOLEPROTEINASE_HIS_UNKNOWN_1.
 SQ SEQUENCE 813 AA; 93499 MW; 55FC07E7329DBCA1 CRC64;
 Query Match 35.5%; Score 50; DB 2; Length 813;
 Best Local Similarity 38.1%; Pred. No. 52;
 Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;
 Oy 4 VLGVLHVEPIESQLGLLS 24
 Db 367 ILGFVHMIPKRRRLDLAS 387
 RESULT 6
 006111 ID 006111 PRELIMINARY; PRT; 275 AA.
 AC 006111:
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
 DE PUTATIVE TRANSPORT SYSTEM PERMEASE PROTEIN.
 OS Clostridium perfringens.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 CC Clostridium.
 OK NCBI_TaxID=1502;
 RP [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NCCT 8239;
 RX MEDLINE=93113001; PubMed=1472712;
 RA Holck A.L., Blom H.;
 RT "The nucleotide sequence of a putative membrane transport gene from
 RT Clostridium perfringens."
 RL DNA Seq. 3:191-194(1992).
 CC -1- FUNCTION: MAY BE PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
 CC SYSTEM.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 CC (POTENTIAL).
 CC -1- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-
 CC PROTEIN-DEPENDENT TRANSPORT SYSTEMS.
 DR EMBL; X66092; CAA46887.1; -
 DR InterPro: IPR000515; -
 DR Pfam: PF00528; BPD_transp. 1.
 DR PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; 1.
 KW Inner membrane; Transport; Transmembrane.
 SQ SEQUENCE 275 AA; 30699 MW; CC6B951258727272 CRC64;

Query Match 34.8%; Score 49; DB 2; Length 275;
 Best Local Similarity 50.0%; Pred. No. 23;
 Matches 10; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
 Oy 8 LHVPEFIESQLGLSPVS 27
 Db 121 IIVPOFLVOKIGLIRTS 140
 RESULT 7
 09VEE9 ID 09VEE9 PRELIMINARY; PRT; 469 AA.
 AC 09VEE9:
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE CG7265 PROTEIN.
 GN CG7265.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OK NCBI_TaxID=7227;
 RP [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 Abirl J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 BA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paule J.M.,
 RA Palazozo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A., Wang X.,
 RA Wang Z.-Y., Wassenaar D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003706; AAF5111.1; -
 DR FlyBase; FBgn038272; CG7265.
 DR InterPro: IPR002728; -
 DR Pfam; PF01866; Diphthamide_syn. 1.
 SQ SEQUENCE 469 AA; 52095 MW; 4D149090874896A2 CRC64;

```

Query Match          34.4%; Score 48.5; DB 5; Length 469;
Best Local Similarity 35.5%; Pred. No. 48;
Matches 11; Conservative 6; Mismatches 11; Indels 3; Gaps 1;

QY 1 RRCV---LCYLVHVPFIESQLGLSPVSL 28
DB 156 QVCVLDIGYQHVGKRLKQSLLEPEKL 186

RESULT 8
Q95309 PRELIMINARY; PRT; 99 AA.
AC Q95309;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DE 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE TEGT (FRAGMENT).
GN TEGT.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RL Winteroe A.K., Fredholm M., Davies W.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; 281159; CAB03546.1;
FT NON_TER 99
SQ SEQUENCE 99 AA; 10846 MW; EA009C94B5EB6211 CRC64;

Query Match          34.0%; Score 48; DB 6; Length 99;
Best Local Similarity 40.9%; Pred. No. 11;
Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 7 YHIVEFIESQLGLSPVSL 28
DB 46 YHVHVFRIQAGLSALGSLGL 67

RESULT 9
Q9VZ03 PRELIMINARY; PRT; 512 AA.
AC Q9VZ03;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CG15812 PROTEIN.
GN CG15812.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Elytrgota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RX Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RX George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RX Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RX Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RX Man K.H., Doyle E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RX Abril J.F., Agdayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RX Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RX Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RX Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RX Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RX Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RX de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RX Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

```

```

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegam C.,
RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Mospreff A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.W., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003477; AAF47724.1;
DR FlyBase; FBgn0035405; CG15812.
DR InterPro; IPR000210;
DR InterPro; IPR002672;
DR Pfam; PF00651; BTB; 1.
DR ProDom; PD010767; -, 1.
DR PROSITE; PS50097; BTB; 1.
DR SMART; SM00225; BTB; 1.
SQ SEQUENCE 512 AA; 57010 MW; E73625C91D4E313 CRC64;

Query Match          34.0%; Score 48; DB 5; Length 512;
Best Local Similarity 31.4%; Pred. No. 63;
Matches 11; Conservative 10; Mismatches 4; Indels 10; Gaps 2;

QY 3 CVLGYLHI-----VPEFIES-QLGLSPVSL 27
DB 74 CMLSFYMGETSLPSASLPFLERINLGIKSAIS 108

RESULT 10
Q9U3X4 PRELIMINARY; PRT; 626 AA.
AC Q9U3X4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE SDHA (EC 1.3.99.1).
GN SDHA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX2;
RC Lay S.P., Fisher P.R.;
RT "The flavoprotein subunit 1 of Dictyostelium succinate
RT dehydrogenase.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -1-COPFACTOR: FAD (BY SIMILARITY).
CC -1-SIMILARITY: TO PYRIDINE NUCLEOTIDE-DISULPHIDE OXIDOREDUCTASES
CC CLASS=1.
DR EMBL; AF211482; AAF21045.1; -.
DR HSSP; P00363; 1FUW.
DR InterPro; IPR000464; -.
DR InterPro; IPR001100; -.
DR InterPro; IPR001327; -.
DR Pfam; PF00890; FAD_binding_2; 1.
DR PRINTS; PR00368; FADPNR.

```


AC Q04919;
AT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE ADRIIP (FRAGMENT).
GN ADRI.
OS *Saccharomyces cerevisiae* (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB972;
RA Murphy L., Harris D.;
RN Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB972;
RA Barrell B., Rajandream M.A.;
RN Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RL [3]
RP SEQUENCE FROM N.A.
RC STRAIN=AB972;
RA MEDLINE=97039860; PubMed=8885406;
RA de Virgilio C., Demarini D.J., Pringle J.R.;
RT "SPR28, a sixth member of the septin gene family in *Saccharomyces cerevisiae* that is expressed specifically in sporulating cells.";
RL Microbiology 142:2897-2905(1996).
DR EMBL: Z48613; CAA88496.1; -.
DR SGD: S0002624; ADRI.
DR SEQUENCE 496 AA; 57177 MW; 5FA9B246B0FCF8CD CRC64;

Query Match	33.3%	Score 47	DB 3	Length 496
Best Local Similarity	46.7%	Pred. No. 86		
Matches	7	Conservative	5	Mismatches 3
				Indels 0
				Gaps 0
Q7	6	GYLHIVPEPIESQL	20	
		: : : : :		
DB	85	GYLHIVPEPIESMT	99	

ID	RESULT	15
027545		
AC	027545	PRELIMINARY;
DT	01-JAN-1998	(TREMBLrel, 05, Created)
DT	01-JAN-1998	(TREMBLrel, 05, Last sequence update)
DT	01-MAR-2001	(TREMBLrel, 16, Last annotation update)
DE	PHENYLALANYL-TRNA SYNTHETASE ALPHA SUBUNIT.	
GN	MTN1501.	
OS	Methanobacterium thermoautotrophicum.	
OC	Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;	
OC	Methanothermobacter.	
OX	NCBI_TaxID=145262;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-DELTA H:	
RA	MEDLINE=98037514; PubMed=9371463.	
RA	Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,	
RA	Alldredge T., Bashlitzadeh R., Blakely D., Cook R., Gilbert K.,	
RA	Harrison D., Hoang L., Keagle P., Lum W., Potthier B., Qiu D.,	
RA	Spadafora R., Vlicare R., Wang Y., Wierzbowski J., Gibson R.,	
RA	Jivani N., Caruso A., Bush D., Safer H., Patweli D., Prabhakar S.,	
RA	McGougal S., Shlmer G., Goyal A., Pietrowski S., Church G.M.,	
RA	Daniels C.J., Mo J.-I., Rice P., Nolling J., Reeve J.N.;	
RT	"Complete genome sequence of Methanobacterium thermoautotrophicum	
RT	delta: functional analysis and comparative genomics."	
RL	J. Bacteriol. 179:7135-7155(1997).	
DR	EMBL: AE000910; AAB85976.1; -	
DR	InterPro: IPR002106; -	
DR	InterPro: IPR002319; -	
DR	Pfam: PF01409; tRNA-synt_2d; 1.	

DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
KW Aminoacyl-tRNA synthetase.
SQ SEQUENCE 532 AA; 61026 MW; 15A011BC4129C95A CRC64;

Query Match	33.3%	Score 47:	DB 1:	Length 533:
Best Local Similarity	35.7%	Pred. No. 93:		
Matches	10:	Conservative	7:	Mismatches 11:
				Indels 0:
				Gaps 0:
OY	1	RTCVGGLHIVPEPTESQLGLTSPVSL	28	
		::: :: :: :: ::		
Db	279	QTEVATRHPLKREWEVATFGSLSPAL	306	

Search completed: August 7, 2001, 13:34:46
Job time: 365 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2001, 13:29:06 ; Search time 8.36 Seconds

(without alignments)
114.731 Million cell updates/sec

Title: US-09-251-133-6

Perfect score: 141

Sequence: 1 RCTVGLYHIVPEFIESQLGLSPVSL 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	34.8	226	1 BIL_RAT	P55062 ratulus norv
2	49	34.8	570	1 ILVD_LACLA	O02139 lactococcus
3	47	33.3	302	1 CFA2_MYCTU	O11136 mycobacteri
4	47	33.3	1323	1 ADRL_YEAST	P07248 saccharomyc
5	46	32.6	183	1 LIPO_BUDMA	O01584 bufo marinu
6	46	32.6	491	1 IRR3_CHICK	O00633 gallus galli
7	46	32.6	4687	1 PLE1_RAT	P30427 rattus norv
8	45	31.9	492	1 SECY_CYAPA	P25014 cyanophora
9	45	31.9	1400	1 RON_HUMAN	O04912 homo sapien
10	44	31.2	237	1 BIL_HUMAN	P55061 homo sapien
11	44	31.2	313	1 NDM_HUMAN	O09824 rhipicephal
12	44	31.2	372	1 CDK9_HUMAN	P50750 homo sapien
13	44	31.2	593	1 SUMT_YEAST	P36130 saccharomyc
14	44	31.2	674	1 PTGA_BRELA	O45298 brevbacter
15	43	30.5	649	1 METX_YEAST	O04533 saccharomyc
16	43	30.5	111	1 YH12_YEAST	P38895 saccharomyc
17	43	30.5	229	1 Y790_MERJA	O58200 methanococ
18	43	30.5	253	1 YMDA_CHLAU	O45826 chlorellaxu
19	43	30.5	326	1 TMOF_PSEME	O03304 pseudomonas
20	43	30.5	435	1 YYS3_CAEEL	P49126 caenorhabdi
21	43	30.5	437	1 CCA_ARCFU	O28126 archaeoglob
22	43	30.5	1131	1 DNBI_HSVJ7	P52339 human herpe
23	42	29.8	272	1 K4YN_RHSN	P55722 rhizobium s
24	42	29.8	360	1 RTCA_AOUAE	O66884 aquilex aeo
25	42	29.8	434	1 FLIT_HELPJ	O07025 helicobacte
26	42	29.8	444	1 FLIT_HELPJ	O07025 helicobacte
27	42	29.8	524	1 MVIN_SALTY	P37169 salmonella
28	42	29.8	662	1 SL51_RABIT	P11170 oryctolagus
29	42	29.8	956	1 HELI_HCMVA	P16736 human cytom
30	42	29.8	1403	1 BIRA_MOUSE	O9qws5 mus musculu
31	41	29.1	82	1 V082_ASFB7	P18557 african swi
32	41	29.1	118	1 V118_ASFB7	P18556 african swi
33	41	29.1	136	1 AT91_BOVIN	P32876 bos taurus

34	41	29.1	136	1 AT91_SHEEP	P17605 ovis aries
35	41	29.1	213	1 RISA_ECOLI	P29015 escherichia
36	41	29.1	214	1 DYS_MYCTU	P31500 mycobacteri
37	41	29.1	290	1 PROB_LEPIN	P94871 leptospira
38	41	29.1	423	1 AIM1_MOUSE	P35855 m adaptor-r
39	41	29.1	511	1 MVIN_ECOLI	P75932 escherichia
40	41	29.1	884	1 RPOB_NPYOP	O12934 oryza pseu
41	41	29.1	893	1 PMTX_SCHPO	O13898 schistosach
42	41	29.1	1118	1 YIPL_YEAST	P40456 saccharomyc
43	40.5	28.7	329	1 GALE_STRLI	P13226 streptomyc
44	40.5	28.7	348	1 EXOB_AZOB	O39083 azospirillum
45	40.5	28.7	391	1 BRB2_HUMAN	P30411 homo sapien

ALIGNMENTS

RESULT	1	STANDARD	PRT	236 AA.
BIL_RAT				
ID	P55062: 064712:			
AC	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	BAX INHIBITOR-1 (BI-1) (TESTIS ENHANCED GENE TRANSCRIPT).			
GN	BIL OR TEST.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_Taxid=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=SPRAGUE-DAWLEY; TISSUE=Testis;			
RX	MEDLINE=94281747; PubMed=8012111;			
RA	Walter L., Dirks B., Rothermel E., Heyens M., Szpirer C., Levan G.,			
RA	Guenther E.,			
RT	"A novel, conserved gene of the rat that is developmentally regulated			
RT	in the testis."			
RL	Mamm. Genome 5:216-221(1994).			
CC	-1- FUNCTION: SUPPRESSOR OF APOPTOSIS (BY SIMILARITY).			
CC	-1- SUBUNIT: INTERACTS WITH BCL2 AND BCL-XL (BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).			
CC	-1- TISSUE SPECIFICITY: HIGHLY ABUNDANT IN TESTIS.			
CC	-1- SIMILARITY: BELONGS TO THE BIL FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: X75855; CA53470.1; -			
DR	EMBL: X75856; CA53471.1; -			
DR	InterPro: IPR002199; -			
DR	Pfam: PF01027; UPF0005.1.			
DR	PROSITE: PS01243; BIL; 1.			
KW	Apoptosis; Transmembrane.			
FT	TRANSMEM 30			POTENTIAL.
FT	TRANSMEM 53			POTENTIAL.
FT	TRANSMEM 88			POTENTIAL.
FT	TRANSMEM 112			POTENTIAL.
FT	TRANSMEM 139			POTENTIAL.
FT	TRANSMEM 166			POTENTIAL.
FT	TRANSMEM 206			POTENTIAL.
FT	TRANSMEM 226			POTENTIAL.
SQ	SEQUENCE 236 AA; 26358 MW; FFA412EC1DC87537 CRC64;			

Query Match 34.8%; Score 49; DB 1; Length 236;
Best Local Similarity 40.9%; Pred. No. 2.4;
Matches 9; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

OY 7 YHVFIEFIESQLGLSPVSL 28
 ID 46 YHVFIEFIESQLGLSALGAL 67

RESULT 2

ILVD_LACLA STANDARD; PRT; 570 AA.

AC 002139;

DT 01-JUL-1993 (rel. 26, Created)

DT 01-JUL-1993 (rel. 26, Last sequence update)

DT 15-DEC-1998 (rel. 37, Last annotation update)

DE DIHYDROXY-ACID DEHYDRATASE (EC 4.2.1.9) (DAD).

GN ILVD.

OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Lactococcus.

OX NCBI_TaxID=1360;

RN NCBI_TaxID=1360;

RP SEQUENCE FROM N.A.

RC STRAIN=NCDO 2118;

RA MEDLINE=93015710; PubMed=1400210;

RA Godon J.-J., Chopin M.-C., Ehrlich D.S.;

RT "Branched-chain amino acid biosynthesis genes in Lactococcus lactis

subsp. lactis."

RL J. Bacteriol. 174:6580-6589(1992).

CC -1- CATALYTIC ACTIVITY: 2,3-DIHYDROXY-3-METHYLBUTANATE = 3-METHYL-2-OXOBUTANATE + H(2)O.

CC -1- COFACTOR: BINDS A 2FE-2S CLUSTER (POTENTIAL).

CC -1- PATHWAY: FOURTH STEP IN VALINE AND ISOLEUCINE BIOSYNTHESIS.

CC -1- SIMILARITY: BELONGS TO THE ILVD / EDD FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation

at the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@sib-sib.ch).

CC -----

DR EMBL: U92974; AAB81918.1.

DR PIR: S35137; S35137.

DR InterPro: IPR000581.

DR Pfam: PF00920; ILVD_EDD.1.

DR PROSITE: PS00886; ILVD_EDD.1; 1.

DR PROSITE: PS00887; ILVD_EDD.2; 1.

KM Branched-chain amino acid biosynthesis: Lyase: Iron-sulfur.

FT METAL 135 135 IRON-SULFUR (2FE-2S) (POTENTIAL).

FT METAL 207 207 IRON-SULFUR (2FE-2S) (POTENTIAL).

SQ SEQUENCE 570 AA; 60718 MW; F4664583A79D0A62 CRC64;

Query Match 34.8%; Score 49; DB 1; Length 570;

Best Local Similarity 60.0%; Pred. No. 6;

Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 9 HIVEFIESQLGL 23

DB 495 HIVEAVEGLGLV 509

RESULT 3

CEFA2_MYCTU STANDARD; PRT; 302 AA.

AC 011196;

DT 01-OCT-1996 (rel. 34, Created)

DT 01-OCT-1996 (rel. 34, Last sequence update)

DT 30-MAY-2000 (rel. 39, Last annotation update)

DE CYCLOPROPANE-FATTY-ACYL-PHOSPHOLIPID SYNTHASE 2 (EC 2.1.1.79)

DE (CYCLOPROPANE FATTY ACID SYNTHASE) (CFA SYNTHASE) (CYCLOPROPANE

MYCOLIC ACID SYNTHASE 2)

GN CMA2 OR RV0503C OR MYCY2069.30C.

OS Mycobacterium tuberculosis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1773;

RN NCBI_TaxID=1773;

RP SEQUENCE FROM N.A.

RC STRAIN=H37RV;

RA MEDLINE=96070840; PubMed=7592990;

RA George K.M., Yuan Y., Sherman D.R., Barry C.E. III;

RT "The biosynthesis of cyclopropanated mycolic acids in Mycobacterium

tuberculosis. Identification and functional analysis of CMA-2.";

RL J. Biol. Chem. 270:27292-27298(1995).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=H37RV;

RA MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

RA Gordon S.V., Eiglsmeier K., Gas S., Barry C.E. III, Tekala F.,

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holtroyd S.,

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J.,

RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

RA Rutter S., Seeger K., Skellern S., Squares S., Squires R., Sulston J.E.,

RA Taylor K., Whitehead S., Barrell B.G.;

RT "Deciphering the biology of Mycobacterium tuberculosis from the

complete genome sequence."

RL Nature 393:537-544(1998).

CC -1- FUNCTION: TRANSFERS A METHYLENE GROUP FROM S-ADENOSYL-L-METHIONINE

TO THE CIS DOUBLE BOND OF AN UNSATURATED FATTY ACID CHAIN

RESULTING IN THE REPLACEMENT OF THE DOUBLE BOND WITH A METHYLENE

BRIDGE. MYCOLIC ACIDS, WHICH REPRESENT THE MAJOR CONSTITUENT OF

MYCOBACTERIAL CELL WALL COMPLEX, ACT AS SUBSTRATES.

CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + PHOSPHOLIPID

CYCLOPROPANE FATTY ACID = S-ADENOSYL-L-HOMOCYSTEINE + PHOSPHOLIPID

CC CYCLOPROPANE FATTY ACID.

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation

at the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@sib-sib.ch).

CC -----

DR EMBL: U34637; AAC34488.1; ALT_INIT.

DR EMBL: Z77162; CAB00929.1.

DR Tuberculist: RV0503C.

KM Transferase: Methyltransferase: Lipid synthase.

FT DOMAIN 76 84 S-ADENOSYL-L-METHIONINE-BINDING MOTIF.

FT ACT_SITE 284 284 ESSENTIAL FOR CATALYSIS (PROBABLE).

SQ SEQUENCE 302 AA; 34660 MW; 63AAA95627F95755 CRC64;

Query Match 33.3%; Score 47; DB 1; Length 302;

Best Local Similarity 55.6%; Pred. No. 6.3;

Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 11 VPEFIESQLGLSPVSL 28

DB 186 IPKKEAQLGLTSPMSL 203

RESULT 4

ADRI_YEAST STANDARD; PRT; 1323 AA.

AC P07248;

DT 01-APR-1988 (rel. 07, Created)

DT 01-APR-1988 (rel. 07, Last sequence update)

DT 15-JUL-1999 (rel. 38, Last annotation update)

DE REGULATORY PROTEIN ADRI.

GN ADRI OR YDR216W OR YD8142.16 OR YD8142B.08.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 RX NCBI_TaxID=4932;
 [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=86175015; PubMed=3515197;
 RT Hartshorne T.A., Blumberg H., Young E.T.;
 RT "Sequence homology of the yeast regulatory protein ADRI with Xenopus
 RT transcription factor TFIID.";
 RL Nature 320:283-287(1986).
 RN [2]
 RP SEQUENCE OF 1-973 FROM N.A.
 RC STRAIN=5288C / AB972;
 RA Oliver K., Shore L., Harris D., Barrell B.G., Rajandream M.A.,
 RA Walsh S.V.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP STRUCTURE BY NMR OF ZINC-FINGERS.
 RX MEDLINE=88336881; PubMed=3047872;
 RA Parraga G., Horvath S.J., Eisen A., Taylor W.E., Hood L., Young E.T.,
 RA Klevit R.E.;
 RT "Zinc-dependent structure of a single-finger domain of yeast ADRI.";
 RL Science 241:1469-1492(1988).
 RN [4]
 RP STRUCTURE BY NMR OF 131-159.
 RX MEDLINE=91190886; PubMed=2012802;
 RA Xu R.X., Horvath S.J., Klevit R.E.;
 RT "ADRI, a zinc finger peptide, exists in two folded conformations.";
 RL Biochemistry 30:3365-3371(1991).
 RN [5]
 RP STRUCTURE BY NMR OF 102-161.
 RX MEDLINE=99260744; PubMed=10331877;
 RA Bowers P.M., Schautler L.E., Klevit R.E.;
 RT "A folding transition and novel zinc finger accessory domain in the
 RT transcription factor ADRI.";
 RL Nat. Struct. Biol. 6:478-485(1999).
 RN [6]
 RP MUTAGENESIS.
 RX MEDLINE=87287210; PubMed=3112579;
 RA Blumberg H., Eisen A., Sledziewski A., Bader D., Young E.T.;
 RT "Two zinc fingers of a yeast regulatory protein shown by genetic
 RT evidence to be essential for its function.";
 RL Nature 328:443-445(1987).
 RN [7]
 RP MUTAGENESIS.
 RX MEDLINE=92021016; PubMed=1924382;
 RA Thukral S.K., Morrison M.L., Young E.T.;
 RT "Alanine scanning site-directed mutagenesis of the zinc fingers of
 RT transcription factor ADRI: residues that contact DNA and that
 RT transactivate.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:9188-9192(1991).
 RN [8]
 RP MUTAGENESIS.
 RX MEDLINE=92269852; PubMed=1588970;
 RA Thukral S.K., Morrison M.L., Young E.T.;
 RT "Mutations in the zinc fingers of ADRI that change the specificity of
 RT DNA binding and transactivation.";
 RL Mol. Cell. Biol. 12:2784-2792(1992).
 RN [9]
 RP MUTAGENESIS.
 RX MEDLINE=92195295; PubMed=1549108;
 RA Denis C.L., Fontaine S.C., Chase D., Kemp B.E., Bemis L.T.;
 RT "ADRIc mutations enhance the ability of ADRI to activate
 RT transcription by a mechanism that is independent of effects on cyclic
 RT AMP-dependent protein kinase phosphorylation of Ser-230.";
 RL Mol. Cell. Biol. 12:1507-1514(1992).
 RN [10]
 RP MUTAGENESIS.
 RX MEDLINE=93078777; PubMed=1448103;
 RA Camier S., Kacherovsky N., Young E.T.;
 RT "A mutation outside the two zinc fingers of ADRI can suppress defects
 RT in either finger.";
 RL Mol. Cell. Biol. 12:5758-5767(1992).
 RN [11]

RP MUTAGENESIS.
 RX MEDLINE=94179364; PubMed=8132676;
 RA Cook W.J., Mosley S.P., Audino D.C., Mullaney D.L., Rovelli A.,
 RA Stewart G., Denis C.L.;
 RT "Mutations in the zinc-finger region of the yeast regulatory protein
 RT ADRI affect both DNA binding and transcriptional activation.";
 RL J. Biol. Chem. 269:9374-9379(1994).
 CC -1- FUNCTION: REQUIRED FOR TRANSCRIPTIONAL ACTIVATION OF GLUCOSE-
 CC REPRESSIBLE ALCOHOL DEHYDROGENASE (ADH2).
 CC -1- SUBCELLULAR LOCATION: NUCLEUS.
 CC -1- PTM: PHOSPHORYLATION AT SER-230 BY CAMP-DEPENDENT PROTEIN KINASE
 CC A DOES NOT AFFECT DNA BINDING BUT APPEARS TO PREVENT TRANSCRIPTION
 CC OF ADH2 DURING GLUCOSE REPRESSION.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: U28414; AAA73863.1; -;
 CC DR EMBL: Z68194; CAA92359.1; -;
 CC DR EMBL: Z68195; CAA92367.1; -;
 CC DR PIR: A24534; A24534.
 CC DR PDB: 1ARD; 31-JAN-94.
 CC DR PDB: 1ARE; 31-JAN-94.
 CC DR PDB: 1PAA; 15-OCT-94.
 CC DR PDB: 2ADR; 17-JUN-98.
 CC DR TRANSFAC: T00011; -;
 CC DR SCD: S0002624; ADRI.
 CC DR InterPro: IPR000822; -;
 CC DR Pfam: PF00096; zf-C2H2; 2.
 CC DR PRINTS: PR00048; ZINC_FINGER_C2H2_1; 2.
 CC DR PROSITE: PS00028; ZINC_FINGER_C2H2_2; 2.
 CC DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 2.
 CC KW Transcription regulation; Activator; DNA-binding; Nuclear protein;
 CC Zinc-finger; Metal-binding; Repeat; Phosphorylation; 3D-structure.
 CC FT DOMAIN 104 155
 CC FT ZN_FING 104 126
 CC FT ZN_FING 132 155
 CC FT MOD_RES 230 230
 CC FT MUTAGEN 106 106
 CC FT MUTAGEN 109 109
 CC FT MUTAGEN 114 114
 CC FT MUTAGEN 118 118
 CC FT MUTAGEN 122 122
 CC FT MUTAGEN 134 134
 CC FT MUTAGEN 142 142
 CC FT STRAND 104 106
 CC FT TURN 107 110
 CC FT STRAND 111 113
 CC FT HELIX 116 126
 CC FT TURN 127 128
 CC SQ SEQUENCE 1323 AA; 150962 MW; 75E4E0DF2A9FAFE CRC64;
 CC -----
 CC Query Match 33.3%; Score 47; DB 1; Length 1323;
 CC Best Local Similarity 46.7%; Pred. No. 30;
 CC Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
 CC QY 6 GYLHIVEFISQIL 20
 CC Db 912 GHLHSIEFGLSSMI 926
 CC -----
 CC RESULT 5
 CC LIPQ_BUFMA STANDARD; PRT; 163 AA.
 CC ID LIPQ_BUFMA
 CC AC 001584;
 CC DT 01-APR-1993 (Rel. 25, Created)

```

DT 01-APR-1993 (rel. 25, last sequence update)
DE 01-NOV-1995 (rel. 32, last annotation update)
DE LIPOCALIN PRECURSOR.
OS Bufo marinus (Giant toad) (Cane toad).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufonidae;
OC Bufo.
OX NCBI_TaxID=8386;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-16; 18-20; 67-73; 86-93;136-141.
RC TISSUE=choroid plexus;
RA MEDLINE=93054646; PubMed=1385415;
RA Achen M.G., Harms P.J., Thomas T., Richardson S.J.,
RA Wettenhall R.E.H., Schreiber G.;
RA "Protein synthesis at the blood-brain barrier. The major protein
RT secreted by amphibian choroid plexus is a lipocalin.";
RL J. Biol. Chem. 267:23170-23174(1992).
CC -1- FUNCTION: MIGHT HAVE A TRANSPORT FUNCTION ACROSS THE BLOOD BRAIN
CC BARRIER. IS SUPPOSED TO HAVE SIMILAR FUNCTIONS AS TRANSTHYRETIN
CC WHICH MUST HAVE EVOLVED AFTER THE STAGE OF THE AMPHIBIANS IN
CC EVOLUTION.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: SECRETED INTO THE CEREBROSPINAL FLUID.
CC -1- TISSUE SPECIFICITY: EXPRESSED MAINLY IN CHOROID PLEXUS. MUCH LOWER
CC EXPRESSION IN OTHER BRAIN AREAS, AND ABSENT FROM LIVER.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT AMPHIBIAN METAMORPHOSIS.
CC -1- SIMILARITY: BELONGS TO THE LIPOCALIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X67952; CAA8138.1; -.
DR EMBL: L06806; AAA8534.1; -.
DR PIR: S25465; S25465.
DR PIR: A44456; A44456.
DR InterPro: IPR000566; -.
DR InterPro: IPR002345; -.
DR InterPro: IPR002972; -.
DR Pfam: PF00061; Lipocalin; 1.
DR PRINTS: PR01179; LIPOCALIN.
DR PRINTS: PR01254; PGDSYNTASE.
DR PROSITE: PS00213; LIPOCALIN; 1.
KW Lipocalin; Signal.
FT SIGNAL 1 20
FT CHAIN 21 183 LIPOCALIN.
FT DISULFID 83 179
SQ SEQUENCE 183 AA; 20608 MW; 79017CDB1BCF2911 CRC64;

Query Match 32.6%; Score 46; DB 1; Length 183;
Best Local Similarity 47.4%; Pred. No. 5.4;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 3 CVLGYLHVPEFIESQLGLG 21
DB 17 CVYGDVPIQDPFDEKILG 35

RESULT 6
ID IRF3_CHICK STANDARD; PRT; 491 AA.
AC Q90643;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 01-NOV-1997 (rel. 35, Last annotation update)
DE INTERFERON REGULATORY FACTOR 3 (IRF-3).
OS Gallus gallus (Chicken).
RX Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA MEDLINE=95334365; PubMed=7541908;
RA Grant C.E., Vasa M.Z., Deeley R.G.;
RT "IRF-3, a new member of the interferon regulatory factor (IRF)
RT family that is rapidly and transiently induced by dsRNA.";
RL Nucleic Acids Res. 23:2137-2146(1995)
CC -1- FUNCTION: MAY ACTIVATE TRANSCRIPTION BY COMPLEX FORMATION WITH
CC OTHER TRANSCRIPTIONAL FACTORS, POSSIBLY MEMBERS OF THE STAT
CC FAMILY. BINDS SPECIFICALLY TO THE IFN-STIMULATED RESPONSE ELEMENT
CC (ISRE) BUT NOT TO THE IRF-1 BINDING SITE PRD-1.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE IRF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U20338; AAA6995.1; -.
DR HSSP: P15314; IIF1.
DR InterPro: IPR001346; -.
DR Pfam: PF00605; IRF; 1.
DR PRINTS: PR00267; INTERNEGCT.
DR PROSITE: PS00601; IRF; 1.
KW Transcription regulation; DNA-binding; Nuclear protein; Activator.
FT DNA_BIND 14 112 TRYPTOPHAN PENTAD REPEAT.
FT SEQUENCE 491 AA; 54441 MW; CAE0C2AA8BE976D9 CRC64;

Query Match 32.6%; Score 46; DB 1; Length 491;
Best Local Similarity 42.3%; Pred. No. 15;
Matches 11; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

OY 3 CVLGYLHVPEFIESQLGLSPVSL 28
DB 307 CVLGYLPDPVAVALRVLPPSPSL 332

RESULT 7
ID PLE1_RAT STANDARD; PRT; 4687 AA.
AC P30427; O08879; O08880; O08881;
DT 01-APR-1993 (rel. 25, Created)
DT 01-OCT-1996 (rel. 34, Last sequence update)
DT 01-OCT-2000 (rel. 40, Last annotation update)
DE PLECTIN 1.
GN PLECT.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=glial tumor;
RC MEDLINE=91268156; PubMed=2050743;
RA Wiche G., Becker B., Luber K., Weitzer G., Castanon M.J.,
RA Hauptmann R., Striowa C., Stewart M.;
RT "Cloning and sequencing of rat plectin indicates a 466-kD polypeptide
RT chain with a three-domain structure based on a central alpha-helical
RT coiled coil.";
RL J. Cell Biol. 114:83-99(1991).
RN [2]
RP REVISIONS.
RC TISSUE=glial tumor;
RC MEDLINE=96210632; PubMed=8633055;

```


CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M30487; AAA63630.1; -
 CC EMBL: U30821; AAA61218.1; -
 DR Mendel; 7931; CYAPA; secy1.1.
 CC InterPro; IPR002208; -
 DR Pfam; PF00344; secy; 1.
 CC PRINTS; PR00303; SECYTRNLCASE.
 DR PROSITE; PS00755; SECY_1; 1.
 CC PROSITE; PS00756; SECY_2; 1.
 KW Protein transport; Transmembrane; Cyanelle; Translocation.
 SQ SEQUENCE 492 AA; 55926 MW; 8736219112BFC0D CRC64;
 Query Match 31.9%; Score 45; DB 1; Length 492;
 Best Local Similarity 41.4%; Pred. No. 22;
 Matches 12; Conservative 5; Mismatches 8; Indels 4; Gaps 1;
 QY 4 VLGYLHIVEPEIESQL-----LGLLSPVSL 28
 DB 419 VLVLCILPSIVSRSLGRLPLRLTLPVSI 447
 RESULT 9
 ID RON_HUMAN STANDARD; PRT; 1400 AA.
 AC 004912;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE MACROPHAGE-STIMULATING PROTEIN RECEPTOR PRECURSOR (EC 2.7.1.112)
 DE (MSP RECEPTOR (P185-RON) (CDW136) (CD136 ANTIGEN)).
 GN MST1R OR RON.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 ON NCBI_Taxid=9606;
 RX [1]
 RP TISSUE=keratinocytes;
 RC MEDLINE=93241719; PubMed=8368824;
 RA Ronsin C., Muscatelli F., Mattei M.-G., Breathnach R.;
 RT "A novel putative receptor protein tyrosine kinase of the met
 family.";
 RL Oncogene 8:1195-1202(1993).
 RN [2]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE=96413302; PubMed=8816464;
 RA Collesi C., Santoro M.M., Gaudino G., Comoglio P.M.;
 RT "A splicing variant of the RON transcript induces constitutive
 tyrosine kinase activity and an invasive phenotype.";
 RL Mol. Cell. Biol. 16:5518-5526(1996).
 RN [3]
 RP FUNCTION.
 RX MEDLINE=95025877; PubMed=7939629;
 RA Wang M.-H., Ronsin C., Gesnel M.-C., Coupey L., Skeel A.,
 RA Leonard E.J., Breathnach R.;
 RT "Identification of the ron gene product as the receptor for the human
 macrophage stimulating protein.";
 RL Science 266:117-119(1994).
 RL CC -1- FUNCTION: RECEPTOR FOR MACROPHAGE STIMULATING PROTEIN (MSP). HAS A
 CC TYROSINE-PROTEIN KINASE ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
 CC PROTEIN TYROSINE PHOSPHATE.
 CC -1- SUBUNIT: HETERODIMER FORMED OF AN ALPHA CHAIN AND A BETA
 CC CHAIN WHICH ARE DISULFIDE LINKED.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: ISOFORM RON (SHOWN HERE) AND
 CC ISOFORM DELTA-RON; ARE PRODUCED BY ALTERNATIVE SPLICING. DELTA-RON
 CC LACKS PART OF THE EXTRACELLULAR DOMAIN, OLIGOMERIZES AND IS
 CC CONSTITUTIVELY ACTIVATED.
 CC -1- TISSUE SPECIFICITY: KERATINOCYTES, AND LUNG.

CC -1- PTM: THE TWO SUBUNITS ARE PROTEOLYTICALLY DERIVED.
 CC -1- PTM: PHOSPHORYLATED IN RESPONSE TO LIGAND BINDING (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO THE MET TYROSINE KINASE FAMILY OF
 CC RECEPTORS.
 CC -1- DATABASE: NAME=PROW; NOTE=CD guide Cw136 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cdw136.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X70040; CAA49634.1; -
 CC PIR: S31736; S31736.
 DR HSSP; P11362; 1FGI.
 DR MIM; 600168; -
 DR InterPro; IPR000719; -
 DR InterPro; IPR001245; -
 DR InterPro; IPR001627; -
 DR InterPro; IPR002165; -
 DR InterPro; IPR002909; -
 DR Pfam; PF01437; Plexin_repeat; 1.
 DR Pfam; PF01403; Sema; 1.
 DR Pfam; PF01833; TIG; 3.
 DR Pfam; PF00069; pkinase; 1.
 DR PRINTS; PR00109; TYRKINASE.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 KW Transferrase; Tyrosine-protein kinase; Receptor; Transmembrane;
 KW Glycoprotein; Signal; Phosphorylation; ATP-binding; Polymorphism;
 KW Alternative splicing.
 FT SIGNL 1 24
 FT CHAIN 25 1400
 FT CHAIN 25 304
 FT CHAIN 310 1400
 FT DOMAIN 25 957
 FT TRANSMEM 958 982
 FT DOMAIN 983 1400
 FT DOMAIN 1082 1345
 FT NP_BIND 1088 1096
 FT BINDING 1114 1114
 FT ACT_SITE 1208 1208
 FT MOD_RES 1353 1353
 FT MOD_RES 1360 1360
 FT CARBOHYD 66 66
 FT CARBOHYD 419 419
 FT CARBOHYD 458 458
 FT CARBOHYD 488 488
 FT CARBOHYD 654 654
 FT CARBOHYD 720 720
 FT CARBOHYD 841 841
 FT CARBOHYD 897 897
 FT VARSPPLIC 932 932
 FT VARIANT 322 322
 FT SEQUENCE 1400 AA; 152227 MW; E8782D07E4D5368A CRC64;
 SQ
 Query Match 31.9%; Score 45; DB 1; Length 1400;
 Best Local Similarity 33.3%; Pred. No. 65;
 Matches 12; Conservative 6; Mismatches 10; Indels 8; Gaps 1;
 QY 1 RUCVLGYLHIVEPEI-----ESQLGLLSPVSL 28
 DB 932 QVCVDECHLGVAVRPGDVGQSTLGLTLPALL 967
 RESULT 10
 ID RON_HUMAN


```

ID      BI1_HUMAN      STANDARD:      PRF:      237 AA.
AC      P55061: 014938:
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      01-OCT-2000 (Rel. 40, Last annotation update)
DE      BAX INHIBITOR-1 (BI-1) (TESTIS ENHANCED GENE TRANSCRIPT).
GN      BI1 OR TEGT.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX      NCBI_TaxID:9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Testis;
RX      MEDLINE=96015061; PubMed=8530040;
RT      Walter L., Marynen P., Szpieler J., Levan G., Guenther E.;
RT      "Identification of a novel conserved human gene, TEGT.";
RL      Genomics 28:301-304(1995).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      Cowling R.T., Birnboim H.C.;
RL      Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN      [3]
RA      CHARACTERIZATION.
RX      MEDLINE=98325348; PubMed=9660918;
RN      Xu Q., Reed J.C.;
RT      "Bax inhibitor-1, a mammalian apoptosis suppressor identified by
RT      functional screening in yeast.";
RT      Mol. Biol. Cell 11:337-346(1998).
CC      -1- FUNCTION: SUPPRESSOR OF APOPTOSIS.
CC      -1- SUBUNIT: INTERACTS WITH BCL2 AND BCL-XL.
CC      -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC      -1- TISSUE SPECIFICITY: HIGHLY ABUNDANT IN TESTIS.
CC      -1- SIMILARITY: BELONGS TO THE BI1 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL; X75861; CAA53472.1; -.
DR      EMBL; AF033095; AAB87479.1; -.
DR      MIM; 600748; -.
DR      InterPro; IPR002199; -.
DR      Pfam; PF01027; UPF0005; 1.
DR      PROSITE; PS01243; BI1; 1.
KW      Apoptosis; Transmembrane.
FT      TRANSMEM 30
FT      TRANSMEM 50
FT      TRANSMEM 73
FT      TRANSMEM 87
FT      TRANSMEM 107
FT      TRANSMEM 113
FT      TRANSMEM 133
FT      TRANSMEM 140
FT      TRANSMEM 160
FT      TRANSMEM 167
FT      TRANSMEM 187
FT      TRANSMEM 207
FT      TRANSMEM 227
FT      CONFLICT 169
FT      CONFLICT 169
FT      CONFLICT 187
FT      CONFLICT 187
FT      SEQUENCE 237 AA; 26473 MW; 120BFB57BC41639E CAC64;
SO      Query Match 31.2%; Score 44; DB 1; Length 237;
      Best Local Similarity 40.9%; Pred. No. 14;
      Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;
OY      7 YLHYPERIESQLGLSPVSL 28
      I::I I::I I::I I::I
DB      46 YVHWYHFIOAGLLSALGSLIL 67

```

```

ID      NUM1_RHISA      STANDARD:      PRT:      313 AA.
AC      099824:
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      01-OCT-2000 (Rel. 40, Last annotation update)
DE      NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 1 (EC 1.6.5.3).
GN      NDI.
OS      Rhinipcephalus sanguineus (Brown dog tick).
OS      Mitochondrion.
OC      Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC      Parasitiformes; Ixodida; Ixodidae; Rhinipcephalus.
OX      NCBI_TaxID=34632.
RX      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=99083443; PubMed=9866211;
RA      Black W.C. IV, Roehndanz R.L.?
RT      "Mitochondrial gene order is not conserved in arthropods: prostriate
RT      and metastriate tick mitochondrial genomes."
RL      Mol. Biol. Evol. 15:1772-1785(1998)
CC      -I- CATALYTIC ACTIVITY: NADH + UBIQUINONE -> NAD(+) + UBIQUINOL.
CC      -I- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb.ch/announce/
CC      or send an email to license@isb.stb.ch).
CC      -----
DR      EMBL, AF081829; AAD05524.1; -.
DR      InterPro; IPR001694; -.
DR      Pfam; PF00146; NADHdh; 1.
DR      PROSITE; PS00667; COMPLEX1_ND1_1; 1.
DR      PROSITE; PS00668; COMPLEX1_ND1_2; 1.
KW      Oxidoreductase; NAD; ubiquinone; Mitochondrion; Transmembrane.
SQ      SEQUENCE 313 AA; 36721 MW; 846DA5BF565EAD1 CRC64;

Query Match      31.2%; Score 44; DB 1; Length 313;
Best Local Similarity 43.5%; Pred. NO. 19;
Matches 10; Conservative 4; Mismatches 9; Indels 0; Gaps 0.

QY      5 LGYHIVPEIESQLGLSPVS 27
DB      30 LGYCHIRKPKNTGAMGLQPI 52
      ||| || : ||| 1:-1
      ||| || : ||| 1:-1

RESULT 12
CDRG9_HUMAN STANDARD: PRT; 372 AA.
AC P50750:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CELL DIVISION PROTEIN KINASE 9 (EC 2.7.1.1-) (SERINE/THREONINE-PROTEIN
DE KINASE PITARE) (C-2K).
GN CDR9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94224836; PubMed=8170997;
RA Grana X., de Luca A., Sang N., Fu Y., Claudio P.P., Rosenblatt J.,
RA Morgan D.O., Giordano A.;
RT "PITARE, a nuclear CD2-related protein kinase that phosphorylates
RT the retinoblastoma protein in vitro."
RL Proc. Natl. Acad. Sci. U.S.A. 91:3834-3838(1994).
RX [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95209665; PubMed=7695608;

```



```

RA STRAIN-ATCC 13869;
RC Yoon K.H.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
CC
CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IID DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL. THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSHO-HPR); IIB TRANSFERS ITS
CC PHOSPHORUS GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC
CC -1- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOHISTIDINE + SUGAR =
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC
CC -1- SIMILARITY: CONTAINS A PTS ETIB DOMAIN.
CC
CC -1- SIMILARITY: CONTAINS A PTS ETIC DOMAIN.
CC
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC its use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-stb.ch/announce/
CC or send an email to license@isb-stb.ch)
CC
CC -----
CC EMBL: L18875; AAA22992.1; -.
CC DR HSSP: P08837; 2F3G.
CC DR InterPro: IPR001127; -.
CC DR InterPro: IPR001996; -.
CC DR Pfam: PF00358; PTS_ETIA.1; 1.
CC DR Pfam: PF00367; PTS_ETIB.1; 1.
CC DR PROSITE: PS00371; PTS_ETIA.1; 1.
CC DR PROSITE: PS01035; PTS_ETIB_CYS.1.
CC KW Phosphotransferase system; Sugar transport; Transferase;
CC KW Phosphorylation; Transmembrane.
CC FT DOMAIN 1 43 ETIB DOMAIN.
CC FT DOMAIN ? ? ETIC DOMAIN.
CC FT DOMAIN 542 674 ETIA DOMAIN.
CC FT TRANSMEM 126 146 POTENTIAL.
CC FT TRANSMEM 162 182 POTENTIAL.
CC FT TRANSMEM 193 213 POTENTIAL.
CC FT TRANSMEM 225 245 POTENTIAL.
CC FT TRANSMEM 260 280 POTENTIAL.
CC FT TRANSMEM 303 323 POTENTIAL.
CC FT TRANSMEM 344 364 POTENTIAL.
CC FT TRANSMEM 377 397 POTENTIAL.
CC FT TRANSMEM 409 429 POTENTIAL.
CC FT TRANSMEM 442 462 POTENTIAL.
CC FT TRANSMEM 493 513 POTENTIAL.
CC FT MOD_RES 28 28 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 594 594 PHOSPHORYLATION (BY SIMILARITY).
CC SQ SEQUENCE 674 AA; 71626 MW; 67A75NAF6EA2FA2 CRC64;

Query Match 31.2%; Score 44; DB 1; Length 674;
Best Local Similarity 60.0%; Pred. No. 43;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 12 PEFISQLGLSPV 26
| | | | | : | : | |
Db 628 PEFIRSKNMLPIITPV 642

RESULT 15
METX_YEAST STANDARD: PRT; 649 AA.
AC Q04533;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PUTATIVE CYSTATINONINE GAMMA-SYNTHASE (EC 4.2.99.9) (O-
DE SUCCINYLHOMOSERINE (THIOU)-LYASE).
IN YML082W.

```

OS *Saccharomyces cerevisiae* (Baker's yeast);
OC Eukaryota; Fungi; Ascomycota; *Saccharomycotina*; *Saccharomycetes*;
OC *Saccharomycetales*; *Saccharomycetaceae*; *Saccharomyces*.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RA Gentiles S., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (NOV-1994) to the EMBL/Genbank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: O-SUCCINYL-L-HOMOSERINE + L-CYSTEINE =
CC CYSATHIONINE + SUCCINATE (CAN ALSO USE HYDROGEN SULFIDE AND
CC METHANETHIOL AS SUBSTRATES).
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- PATHWAY: SECOND STEP IN METHIONINE BIOSYNTHESIS.
CC -1- SIMILARITY: STRONG, TO N.CRASSA MET-7 AND TO YEAST YJR130C.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z46660; CA868656.1; -;
DR SCD; S0004547; YML082W.
DR InterPro; IPR000277; -;
DR Pfam; PF01053; Cys_Met_Meta.P; 1.
KW Hypothetical protein; Methionine biosynthesis; Lyase;
KW pyridoxal phosphate.
FT BINDING 451
FT SEQUENCE 649 AA; 74313 MW; FA952BBA0500BF6E CCK64; PYRIDOXAL PHOSPHATE (POTENTIAL).

Query Match	30.9%	Score 43.5	DB 1	Length 649
Best Local Similarity	34.6%	Pred. NO. 49		
Matches	9	Conservative	6	Mismatches
			6	Indels
			5	Gaps
QY	1	RTCVAGT-----LRIIVPEFIESQLLG	21	
	:	:	:	:
	:	:	:	:
	:	:	:	:
Db	347	KTIVGFPYADTLHVLQSENFETFLG	372	

Search completed: August 7, 2001, 13:35:07
Job time: 361 sec

RESULT	15
METX_YEAST	
ID	METX_YEAST
STANDARD;	PRT;
	649 AA.
AC	Q04533;
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	01-NOV-1997 (Rel. 35, Last annotation update)
DE	PUTATIVE CYSTATINONE GAMMA-SYNTASE (EC 4.2.99.9) (O-SUCCINYLMOMOSERINE (THIOL)-LYASE).
DE	YML082W.

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2001, 13:25:16 ; Search time 15.4 Seconds

(without alignments)
138.499 Million cell updates/sec

Title: US-09-251-133-6

Perfect score: 141

Sequence: 1 RTCVLGYLHIVPEFIESQLGLSPVSL 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.68:*
2: PIR1:*
3: PIR2:*
4: PIR3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	37.6	931	2 H86387	hypothetical prote
2	50	35.5	813	2 A72203	cellobiose-phospho
3	49	34.8	236	2 S42069	TEGT protein - rat
4	49	34.8	275	2 A56641	probable membrane
5	49	34.8	570	2 S35137	probable phosphogl
6	49	34.8	570	2 G86777	dihydroxy-acid deh
7	48	34.0	702	2 E72775	probable helicase
8	47	33.3	302	2 B70746	probable cma22 pro
9	47	33.3	308	2 S72886	hypothetical prote
10	47	33.3	532	2 C69067	phenylalanine--trn
11	47	33.3	656	2 B86033	hypothetical prote
12	47	33.3	1323	2 A24534	regulatory protein
13	46.5	33.0	554	2 T18858	hypothetical prote
14	46	32.6	159	2 T08974	hypothetical prote
15	46	32.6	179	2 H71367	conserved hypochet
16	46	32.6	183	2 S25465	interferon regulat
17	46	32.6	491	2 S56753	hypothetical prote
18	46	32.6	662	2 D86320	hypothetical prote
19	46	32.6	4687	1 A39638	plectin - rat
20	45.5	32.3	616	2 T00894	hypothetical prote
21	45.5	32.3	643	2 D86167	protein F21B7.27 l
22	45.5	32.3	871	2 H72597	hypothetical prote
23	45.5	32.3	982	2 S00954	pol polypeptide -
24	45.5	32.3	2048	2 C84609	hypothetical prote
25	45	31.9	310	2 T30674	hypothetical prote
26	45	31.9	388	2 F70430	hypothetical prote
27	45	31.9	432	2 E96880	hypothetical prote
28	45	31.9	477	2 D82390	Na+/H+ antiporter
29	45	31.9	492	2 T06875	preprotein translo

30	45	31.9	1319	2 H84542	hypothetical prote
31	45	31.9	1400	1 I38185	protein-tyrosine k
32	44.5	31.6	240	2 E69004	hypothetical prote
33	44.5	31.6	648	2 T23864	hypothetical prote
34	44	31.2	237	2 T38334	TEGT (testis enhan
35	44	31.2	313	2 T11160	NADH dehydrogenase
36	44	31.2	317	2 C83652	hypothetical prote
37	44	31.2	344	2 H71825	probable histidine
38	44	31.2	372	2 A55262	protein kinase (EC
39	44	31.2	593	2 S38145	uroporphyrinogen m
40	44	31.2	612	2 T05331	hypothetical prote
41	44	31.2	648	2 T41915	hypothetical prote
42	43.5	30.9	420	2 F69144	O-antigen transpor
43	43.5	30.9	649	2 S49644	hypothetical prote
44	43.5	30.9	650	2 T38692	probable serine/th
45	43.5	30.9	1169	2 T30207	dynein heavy chain

ALIGNMENTS

```

RESULT 1
H86387
hypothetical protein AAC29216.1 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H86387
R:Rheologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Greasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maltl, R.; Marzla
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: H86387
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-931 <STO>
A:Cross-references: GB:AE005172; NID:g11079505; PIDN:AAC29216.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 37.6%; Score 53; DB 2; Length 931;
Best Local Similarity 37.5%; Pred. No. 6.5;
Matches 9; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

OY 3 CVLGYLHIVPEFIESQLGLSPV 26
Db 246 CINOYVONFPELIESQLGLSPV 269

RESULT 2
A72203
cellobiose-phosphorylase - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: A72203
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.W.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316
A:Accession: A72203
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-813 <ARN>
A:Cross-references: GB:AE001822; GB:AE000512; NID:g4982429; PIDN:AAD36910.1; PID:g498

```

A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM1848

Query	Match	Similarity	Score	DB	Length
367	Best Local	38.1%	Pred. No. 16;	813;	
QY	Matches	8;	Conservative	7;	Mismatches
				6;	Indels
				0;	Gaps
QY	4	VLGYHLIVPEFTESQLTGILS	24		
		: : : :			
Db	367	ILGFVHMIPKARORITDLAS	387		
		: : : :			

RESULT 3
S42069 TEGT protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 21-Jan-2000
C:Accession: S42069; I57015; I76675
R:Guenther, E.
submitted to the EMBL Data Library, January 1994
A:Reference number: S42069
A:Accession: S42069
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-236 <GDU>
A:Cross-references: EMBL:X75855; NID:9456206; PIDN:CAA53470.1; PID:9456207
R:Walter, L.; Dirs, B.; Rothermel, E.; Heyens, M.; Szpirer, C.; Levan, G.; Gunther, E.
Mamm. Genome 5, 216-221, 1994
A:Title: A novel, conserved gene of the rat that is developmentally regulated in the testis
A:Reference number: I57015; MUID:94281747
A:Accession: I57015
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-236 <RES>
A:Cross-references: EMBL:X75855; NID:9456206; PIDN:CAA53470.1; PID:9456207
A:Accession: I76675
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 2-236 <REZ>
A:Cross-references: EMBL:X75856; NID:9456208; PIDN:CAA53471.1; PID:9456209
C:Genetics:
A:Gene: Tegt
A:Superfamily: human testicular protein TEGT

Query Match	34.8%	Score 49	DB 2	Length 236
Best Local Similarity	40.9%	Pred. No. 6		
Matches	9	Conservative	6	Mismatches 7; Indels 0; Gaps 0;
QY	7	YLHIVFEFIESOLLGLSPVSL	28	
		: : : :		
db	46	YHVHVRFRIOAGLLSALGAL	67	

```

RESULT      4
A:56641
Probable membrane transport protein - Clostridium perfringens
C:Species: Clostridium perfringens
C:Date: 11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 20-Aug-1999
C:Accession: A56641
R:Holck, A. L.; Blom, H.
DNA Seq. 3, 191-194, 1992
A:Title: The nucleotide sequence of a putative membrane transport gene from Clostridium
A:Reference number: A56641; MUID:93113001
A:Accession: A56641
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-275 <HOL>
A:Cross-references: GH:X66092; NID:g296355; PIDN:CA46887.1; PID:g296356
A:Note: nucleotide sequence not given; conceptual translation not complete
A:Superfamily: maltose transport protein malG

```

Query Match	34.8%	Score 49	DB 2	Length 275
Similarity	50.0%	Pred. No	7.1	
Local				
Matches 10, Conservative	5	Mismatches	5	Indels 0
Gaps	0			
Qy	8	LHIVEFIESQLGLTSPVS	27	
		: : : : : : : : :		
Db	121	IFIVPQFLVOKTGLINTIS	140	

RESULT 5
S35137
probable phosphogluconate dehydratase (EC 4.2.1.12) - *Lactococcus lactis* subsp. *lactis*
C:Species: *Lactococcus lactis* subsp. *lactis*
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 21-Jul-2000
C:Accession: S35137
R:Gordon, J.J.; Chopin, M.C.; Ehrlich, S.D.
J. Bacteriol. 174, 6580-6589, 1992
A:Title: Branched-chain amino acid biosynthesis genes in *Lactococcus lactis* subsp. *lactis*
A:Reference number: S35132; MUID:93015710
A:Accession: S35137
A:Molecule type: DNA
A:Residues: 1-570 <GDD>
A:Cross-references: EMBL:M90761; NID:g2565137; PIDN:AMB81918.1; PID:g2565156
C:Genetics:
A:Gene: *llvD*
C:Superfamily: dihydroxy-acid dehydratase
C:Keywords: branched-chain amino acid biosynthesis; carbon-oxygen lyase; hydro-lyase

Query Match	34.8%	Score 49	DB 2	Length 570
Best Local Similarity	60.0%	Pred. No. 16		
Matches 9	Conservative 3	Mismatches 3	Indels 0	Gaps 0
OY	9	HIVPEFIESQLGLL	23	
Db	495	HIVPEAVEGGLGLV	509	

RESULT 6
G86777
dihydroxy-acid dehydratase (EC 4.2.1.9) [imported] - *Lactococcus lactis* subsp. *lactis*
C:Species: *Lactococcus lactis* subsp. *lactis*
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 31-Mar-2001
C:Accession: G86777
R:Boilotin, A.; Wincker, P.; Mauger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehr
Genome Res. In press, 2001 .
A:Title: The complete genome sequence of the lactic acid bacterium.
A:Reference number: AB6625
A:Accession: G86777
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1*570 <STO>
A:Cross-references: GB:AE005176; NID:q12724193; PIDN:AAK05321.1; GSPDB:GN00146
A:Experimental source: strain IL1403

Query Match	34.8%	Score 49	DB 2	Length 570	
Best Local Similarity	60.0%	Pred. No. 16			
Matches	3	Conservative	3	Mismatches	0
QY	9	HIVPEFISQLGLL	23		
			1:11		
DB	495	HIVPEAVEGLGLV	509		

RESULT

Probable helicase APE0191 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 17-Nov-2000

C:Accession: E72775

R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K

DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr

A:Reference number: A72450; MUID:99310339

A:Accession: E72775

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-702 <KAW>

A:Cross-references: DDBJ:AP000058; NID:95103388; PIDN:BAA79103.1; PID:41042879; PID:9510

A:Experimental source: strain K1

C:Genetics:

A:Gene: APE0191

Query Match 34.0%; Score 48; DB 2; Length 702;

Best Local Similarity 47.4%; Pred. No. 28;

Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Y 4 VLGYLHIVPEIESQLGL 22

Db 372 ISGYRAPPERVESRLACL 390

RESULT 8

Probable cmaA2 protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: B70746

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

Rajandream, M.A.; Rogers, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtrold, S.

Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987

A:Accession: B70746

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-302 <COL>

A:Cross-references: GB:Z77162; GB:AL123456; NID:93261606; PIDN:CAB00929.1; PID:e255156;

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: cmaA2

Query Match 33.3%; Score 47; DB 2; Length 302;

Best Local Similarity 55.6%; Pred. No. 16;

Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Y 11 VPEIESQLGLSPVSL 28

Db 186 IPDKERAGELGUTPMSL 203

RESULT 9

hypothetical protein B2168_F3_130 - Mycobacterium leprae

C:Species: Mycobacterium leprae

C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001

C:Accession: S72886

R:Smith, D.R.; Robison, K.

submitted to the EMBL Data Library, November 1993

A:Description: Mycobacterium leprae cosmid B2168.

A:Reference number: S72586

A:Accession: S72886

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-308 <SMI>

A:Cross-references: EMBL:U00018; NID:9467037; PIDN:AA17222.1; PID:9467038

Query Match 33.3%; Score 47; DB 2; Length 308;

Best Local Similarity 52.2%; Pred. No. 16;

Matches 12; Conservative 5; Mismatches 4; Indels 2; Gaps 1;

Y 8 LR-IVPEIESQLGLSPVSL 28

Db 187 LHTIVPAKEREKELGUTPMSL 209

RESULT 10

phenylalanine--tRNA ligase (EC 6.1.1.20) alpha chain - Methanobacterium thermoautotro

C:Species: Methanobacterium thermoautotrophicum

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999

C:Accession: C69067

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.

Qi, D.; Spadafora, R.; Vicalre, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jivanl,

ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu

A:Reference number: A69000; MUID:98037514

A:Accession: C69067

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-532 <MTH>

A:Cross-references: GB:AE000910; GB:AE000666; NID:92622610; PIDN:AAB85976.1; PID:9262

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTH501

A:Start codon: TTG

C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 33.3%; Score 47; DB 2; Length 532;

Best Local Similarity 35.7%; Pred. No. 30;

Matches 10; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

Y 1 RTCVGLHIVPEIESQLGLSPVSL 28

Db 279 QTEVAYHPLKKEWVATFGLXPISAL 306

RESULT 11

hypothetical protein Z5002 [imported] - Escherichia coli (strain O157:H7)

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001

C:Accession: B86033

C:Perena, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

iller, L.; Grobeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamousis, K.; Apoda

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85460; MUID:21074935; PMID:11206551

A:Accession: B86033

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-656 <STO>

A:Cross-references: GB:AE005174; NID:912518318; PIDN:AAG58726.1; GSPDB:GN00145; UWGP:

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: Z5002

C:Superfamily: Thermotoga maritima hypothetical protein TM0280

Query Match 33.3%; Score 47; DB 2; Length 656;

Best Local Similarity 47.6%; Pred. No. 37;

Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

```

Oy      6 GYLHIVPEFIESQLGLSPV 26
          ||||| |: | | |:
Db     512 GYLHITREMQEGDTLNLTPM 53

```

RESULT 12
A24534
regulatory protein ADRI - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: protein YD8142B.08; protein YDR216W
C:Species: *Saccharomyces cerevisiae*
C:Date: 20-Aug-1987 #sequence.revision 20-Aug-1987 #text_change 06-Feb-1998
C:Accession: A24534; S59423; S59739; S61583
R:Ratshorne, T.A.; Blumberg, H.; Young, E.T.
Nature 320, 283-287, 1986
A:Title: Sequence homology of the yeast regulatory protein ADRI with Xenopus transcript
X:Accession: A24534; M01D:66175015
X:Accession: A24534

[illegible]

```

RESULT      13
             T18858
hypothetical protein C0ZC6.2 - Caenorhabditis elegans
C/S:Species: Caenorhabditis elegans
C/S:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C/S:Accession: T18858
R:Swainburne, J.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19032
A:Accession: T18858
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-554 <MIL>

```

A:Cross-references: EMBL:Z79596; PTDN:CA01856.2; GSPDB:G000028; CESP:C02C6.2
A:Experimental source: clone C02C6
C:Genetics:
A:Gene: CESP:C02C6.2
A:Map position: X
A:Mutons: 13/2, 126/3, 386/2, 419/2, 469/3, 525/3
C:Superfamily: Caenorhabditis elegans hypothetical protein C02C6.2

```

Query Match      33.0%; Score 46.5; DB 2; Length 554;
Best Local Similarity 38.2%; Pred. No. 37;
Matches 13; Conservative 3; Mismatches 5; Indels 13; Gaps 2

QY    1 RTCYL-----GYHVP---EFISQLG 21
      |||         | : | |||| : ||
Db    183 RTCOLLRIECHOVPAHLTIPALEFIEEMIG 216
```

RESULT 14
T08974
hypothetical protein F6G3.10 - *Arabidopsis thaliana*
C:Species: *Arabidopsis thaliana* (mouse-ear cress)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 18-Aug-2000
C:Accession: T08974
R:Bayan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.
submitted to the Protein Sequence Database, May 1999
A:Reference number: Z16520
A:Accession: T08974
A:Molecule type: DNA
A:Residues: 1-169 <BEV>
A:Cross-references: EMBL:AL078464; GSPDB:GN00062; ATSP:F6G3.10
A:Experimental source: cultivar Columbia; BMC clone F6G3
C:Genetics:
A:Gene: ATSP:F6G3.10
A:Map position: 4
A:Superfamily: *Arabidopsis thaliana* hypothetical protein F25G13.40

Query Match	32.6%	Score 46	DB 2	Length 169
Best Local Similarity	36.8%	Pred. No. 12		
Matches	7	Conservative	7	Mismatches 5; Indels 0; Gaps 0
QY	9	HIVPERIESQLGGLSPVS	27	
		: : : : :	:	
Db	93	HIVPQWLDFSVISLIMPF	111	

RESULT 15
H71367
conserved hypothetical protein TP0087 - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C:Date: 24-Jul-1998 #sequence_rev:10n 24-Jul-1998 #text_change 05-Nov-1999
C:Accession: H71367
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G
rison, J.; Khatalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uitterback, T.; M
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:9832770
A:Accession: H71367
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-179 <COL>
A:Cross-references: GB:AE001193; GB:AE000520; NID:g3322343; PIDN:MAC65082.1; PID:g332
A:Experimental source: strain Nichols
C:Genetics:
C:Gene: TP0087

Query Match	32.68;	Score 46;	DB 2;	Length 179;
Best Local Similarity	52.48;	Pred. No. 13;		
Matches 11;	Conservative 2;	Mismatches 8;	Indels 0;	Gaps 0

QY 7 YLHIVPEIESQLGLSPVS 27
| | | | : | | | | :
Db 12 YLHRTPAALSLILGLSCVA 32

Search completed: August 7, 2001, 13:29:28
Job time: 252 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 7, 2001, 13:22:45 ; Search time 19.06 Seconds
(without alignments)
89.059 Million cell updates/sec

Title: US-09-251-133-6

Perfect score: 141
Sequence: 1 RTCVLGYLHIVEPIESQLGLSPVSL 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_0601:*

1: /SIDSR/gcgdata/geneseq/geneseq/AA1980.DAT:*

2: /SIDSR/gcgdata/geneseq/geneseq/AA1981.DAT:*

3: /SIDSR/gcgdata/geneseq/geneseq/AA1982.DAT:*

4: /SIDSR/gcgdata/geneseq/geneseq/AA1983.DAT:*

5: /SIDSR/gcgdata/geneseq/geneseq/AA1984.DAT:*

6: /SIDSR/gcgdata/geneseq/geneseq/AA1985.DAT:*

7: /SIDSR/gcgdata/geneseq/geneseq/AA1986.DAT:*

8: /SIDSR/gcgdata/geneseq/geneseq/AA1987.DAT:*

9: /SIDSR/gcgdata/geneseq/geneseq/AA1988.DAT:*

10: /SIDSR/gcgdata/geneseq/geneseq/AA1989.DAT:*

11: /SIDSR/gcgdata/geneseq/geneseq/AA1990.DAT:*

12: /SIDSR/gcgdata/geneseq/geneseq/AA1991.DAT:*

13: /SIDSR/gcgdata/geneseq/geneseq/AA1992.DAT:*

14: /SIDSR/gcgdata/geneseq/geneseq/AA1993.DAT:*

15: /SIDSR/gcgdata/geneseq/geneseq/AA1994.DAT:*

16: /SIDSR/gcgdata/geneseq/geneseq/AA1995.DAT:*

17: /SIDSR/gcgdata/geneseq/geneseq/AA1996.DAT:*

18: /SIDSR/gcgdata/geneseq/geneseq/AA1997.DAT:*

19: /SIDSR/gcgdata/geneseq/geneseq/AA1998.DAT:*

20: /SIDSR/gcgdata/geneseq/geneseq/AA1999.DAT:*

21: /SIDSR/gcgdata/geneseq/geneseq/AA2000.DAT:*

22: /SIDSR/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	36.2	552	22	AA848241
2	49.5	35.1	188	19	AAW76533
3	49	34.8	236	22	AA87615
4	49	34.8	570	15	AA854219
5	47.5	33.7	171	18	AAV11333
6	47.5	33.7	173	19	AAV38569
7	46.5	33.0	173	19	AA80687
8	46	32.6	127	21	AA824844
9	46	32.6	169	21	AA824843
10	46	32.6	181	21	AA824842
11	46	32.6	183	22	AA890659

12	45.5	32.3	125	18	AAW28030	Staphylococcus aur
13	45	31.9	296	21	AA824612	Arabidopsis thalia
14	45	31.9	328	21	AA824611	Arabidopsis thalia
15	45	31.9	354	21	AA824610	Arabidopsis thalia
16	45	31.9	822	20	AAW73486	Cellobiose phospho
17	45	31.9	1400	20	AAW82791	Human RON receptor
18	44	31.2	122	21	AA803753	Human secreted pro
19	44	31.2	139	22	AA861120	Human herpesvirus
20	44	31.2	237	19	AAW73136	Bax inhibitor BI-1
21	44	31.2	255	21	AA858178	Lung cancer associ
22	44	31.2	372	17	AAW04869	Cyclin-dependent k
23	44	31.2	372	17	AAW04871	Phosphorylation de
24	44	31.2	374	21	AAW8375	Lung cancer associ
25	44	31.2	648	20	AAW6198	Human herpesvirus-
26	44	31.2	648	20	AAW6199	Human herpesvirus-
27	44	31.2	2595	20	AAV39297	Spn a polyketide
28	43.5	30.9	160	21	AAW39660	Human membrane-ass
29	43.5	30.9	1098	22	AAW6533	Corynebacterium gl
30	43.5	30.9	1874	22	AAW6532	Corynebacterium gl
31	43	30.5	286	20	AAV35318	Chlamydia pneumoni
32	43	30.5	308	22	AAW36408	Secreted protein k
33	43	30.5	326	22	AAW60228	Pseudomonas mendoc
34	43	30.5	378	22	AAW65658	Novel protein kina
35	43	30.5	435	21	AAW00199	Putative polyunsat
36	43	30.5	501	21	AAW54157	Human pancreatic c
37	43	30.5	535	19	AAW77299	Amino acid sequenc
38	43	30.5	535	22	AAW65657	Novel protein kina
39	42.5	30.1	372	22	AAW72816	Calendula officina
40	42.5	30.1	665	21	AAW42561	Human ORF2325
41	42	29.8	69	21	AAW44200	Human cancer assoc
42	42	29.8	259	20	AAV35625	Chlamydia pneumoni
43	42	29.8	269	20	AAV08563	B. subtilis hydrol
44	42	29.8	394	19	AAW49681	Open reading frame
45	42	29.8	459	21	AAW24717	Arabidopsis thalia

ALIGNMENTS

RESULT	1
AA848241	standard; Protein: 552 AA.
XX	XX
AA848241:	
AC	XX
XX	XX
02-APR-2001	(first entry)
XX	XX
DE	Amino acid sequence of bzIP2 ORF1 protein.
XX	XX
KW	Transcription factor: seed storage protein: lectin: oil-body protein:
KW	Pv-Seed factor-1; ROM1; VICilin-box binding protein-1; ROM2; 7S-globulin:
KW	phaseolin; PHA-L; bean; nuclear protein; promoter; ORF; bzIP;
KW	basic leucine zipper.
OS	Phaseolus vulgaris.
XX	XX
Key	Location/Qualifiers
FT	Misc-difference 1..552 /note= "Xaa are residues encoded by internal stop codons"
XX	XX
PN	US6160202-A.
XX	XX
PD	12-DEC-2000.
XX	XX
PF	06-FEB-1997; 97US-0796899.
XX	XX
PR	07-OCT-1994; 94US-0319544.
XX	XX
PA	(UYMA-) UNIV MARYLAND BALTIMORE COUNTY.
XX	XX
PI	Chern M, Bustos MM;
XX	XX
DR	WPI; 2001-079619/09.

DR	N-PSDB; AAC84565.
XX	
PT	Novel transcription factor gene which encodes transcription factor
PT	protein that targets promoters of genes encoding seed storage proteins
PT	are useful for modulating seed storage protein expression in dicot seed
PT	crops -
XX	
XX	
PS	Disclosure; Columns 31-36; 67pp; English.
CC	
CC	The invention relates to an isolated transcription factor gene which is
CC	expressed in a recombinant maturing dicot seed and which encodes a
CC	transcription factor protein which targets a promoter of a gene encoding
CC	seed storage proteins, lectins or oil-body proteins. The transcription
CC	factors isolated are Pv-Seed factor-1 (ROM1) and Vicilin-box binding
CC	protein-1 (ROM2). These factors bind to 7S-globulin (b-phaseolin) or
CC	lectin (PNA-) promoters. The transcription factor gene is useful for
CC	enhancing or reducing expression of seed storage protein, lectin or
CC	oil-protein genes in dicot seed crops. The present sequence represents
CC	the amino acid sequence of bzr1p2 (basic leucine zipper) ORF1 protein.
XX	
SQ	Sequence 552 AA;
Query Match	36.2%; Score 51; DB 22; Length 552;
Best Local Similarity	55.0%; Pred. No. 6.7;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0.	
DY	4 VLGYLHIVPEFTESQLGLL 23 ::: Db 121 llgfhltvkefphlllgl 140
RESULT 2	
AAM76633	
ID AAM76633 standard; Protein; 188 AA.	
XX	
AC AAM76633;	
XX	
DE 12-JUL-1999 (first entry)	
XX	
Human herpes virus type 8 FLIP ORF 71 protein.	
XX	
KW Death effector domain; human; murine; anti-apoptotic; treatment;	
KM HIV Infection; autoimmune disease.	
OS Human herpes virus.	
XX	
PN DE19713393-A1.	
PD 08-OCT-1998.	
XX	
PJ 01-APR-1997; 97DE-1013393.	
XX	
PR 01-APR-1997; 97DE-1013393.	
PA (TSCH/) TSCHOOP J.	
PA (ABOT-) APOTECH SA.	
PI Bodmer J, Burns K, French ET, Hahne M, Hoffmann K;	
PI Immler M, Rimoldi D, Schneider P, Schroeter M, Steiner V;	
PI Thome M, Tschopp J, Hofmann K;	
XX	
DR WPI; 1998-532710/46.	
XX	
PT New DNA encoding for anti-apoptotic gene product - used to treat HIV	
PT infections and autoimmune diseases	
PS Claim 16; Fig 16; 45pp; German.	
CC This invention describes novel human and mouse anti-apoptotic gene	
CC products which contain at least one death effector domain. The products	
CC of the invention are used in the treatment of HIV infections and	
CC autoimmune diseases. This sequence represents the human Herpes virus	

```

CC type 8 ORF 71 FLIP protein which is used in the method of the invention.
XX
SQ Sequence 188 AA;

Query Match 35.1%; Score 49.5; DB 19; Length 188;
Best Local Similarity 44.0%; Pred. No. 3.4;
Matches 11; Conservative 4; Mismatches 9; Indels 1; Gaps 1;

OY 1 RTCVL-GYLHIPEFIESQLGLLS 24
   | | | | | | | | | | | | | |
DB 65 rrcilrdllhldprflierhlagtms 89

RESULT 3
AAB87615
ID AAB87615 standard; protein; 236 AA.
XX
AC AAB87615;
XX
DT 15-MAY-2001 (first entry)
XX
DE Bovine mammary tissue derived protein #6.
XX
KW Bovine; mammary gland; cancer; tumour; angiogenesis.
XX
OS Bos taurus.
XX
PN WO200114553-A1.
XX
PD 01-MAR-2001.
XX
PF 23-AUG-2000; 2000WO-NZ00166.
XX
PR 23-AUG-1999; 99US-0150330.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (NZPA-) NEW ZEALAND PASTORAL AGRIC RES INST LTD.
PI Havukkala IJ, Gleen M, Grigor MR, Molenaar AJ;
PI
XX
DR WPI, 2001-226619/23.
XX
XX
PT New polypeptides and polynucleotides encoding the polypeptides, which
PT are expressed in bovine mammary gland tissue, useful for stimulating
PT mammary gland growth or function, or inducing differentiation of milk
PT producing cells -
XX
PS Claim 11; Page 62; 97pp; English.
XX
XX
CC The present invention relates to proteins derived from bovine
CC mammary gland cells. The invention is useful for stimulating
CC bovine mammary gland cell growth and function, inhibiting the
CC growth of various mammary gland cancer cells, inhibiting
CC angiogenesis and vascularization of tumours, or modulating
CC the growth of blood vessels in a mammal.
XX
SQ Sequence 236 AA;

Query Match 34.8%; Score 49; DB 22; Length 236;
Best Local Similarity 40.9%; Pred. No. 5.3;
Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

OY 7 YLHIPEFIESQLGLLSPVSL 28
   | | | | | | | | | | | | | |
DB 46 ylnhvthfaglsaisglsigl 67

RESULT 4
AAR54219
ID AAR54219 standard; Protein; 570 AA.
XX

```

AC AAR54219;
 XX
 DT 09-NOV-1994 (first entry)
 XX
 DE L.lactis branched amino acid synthesis llyd gene product.
 XX
 KW branched amino acid; lly operon; leucine; isoleucine; valine;
 KM blosynthesis; alpha-acetolactate synthase; diacetyl; food flavouring;
 XX attenuation; anti-terminator; Lactococcus.
 OS Lactococcus lactis (subsp. lactis).
 XX
 XX FR2696190-A.
 XX
 PD 01-APR-1994.
 XX
 PF 25-SEP-1992; 92FR-0011470.
 XX
 PR 25-SEP-1992; 92FR-0011470.
 XX
 PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
 XX (AGRI-) AGRIC & FOOD RES COUNCIL.
 XX
 PI Ehrlich S, Godon J, Renault P;
 XX
 DR WPI: 1994-128287/16.
 DR N-PSDB; AAO64211.
 XX
 PT DNA coding for alpha-aceto:lactate synthase - for enhancing
 PT di:acetyl prodn. in microorganisms, esp. for mfr. of dairy prods.
 XX
 PS Disclosure; Flg 2; 45pp; French.
 XX
 CC The genes involved in the pathway for synthesis of branched amino
 CC acids in L.lactis subsp. lactis are organised in two units
 CC containing the leu and lly (including llyd) genes, respectively.
 CC Both units are necessary for the synthesis of leucine but only the
 CC second unit is required for synthesis of ile and Val. The llyb
 CC and llyv genes and the subunits of alpha-acetolactate synthase
 CC that they code for are claimed.
 XX
 SQ Sequence 570 AA;

Query Match 34.8%; Score 49; DB 15; Length 570;
 Best Local Similarity 60.0%; Pred. No. 15;
 Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 9 HIVEFIESQLGL 23
 ||||:|:|:|:
 Db 495 hlypeavegglly 509

RESULT 5
 ID AAY11333 standard; Protein; 171 AA.
 XX
 AC AAY11333;
 XX
 DT 20-MAY-1999 (first entry)
 XX
 DE S. pneumoniae protein SEQ ID NO:443.
 XX
 KM Streptococcus pneumoniae strain 0100993; vaccine; immune response;
 KM streptococcal infection; pneumococcal.
 XX
 OS Streptococcus pneumoniae.
 XX
 XX WO9737026-A1.
 XX
 PD 09-OCT-1997.
 XX
 PF 01-APR-1997; 97WO-US05306.

XX
 PR 22-AUG-1996; 96US-0025788.
 PR 02-APR-1996; 96US-0014690.
 XX
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 XX (SMIK) SMITHKLINE BEECHAM PLC.
 XX
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
 PI Stodola RK;
 XX
 DR WPI: 1997-503111/46.
 DR N-PSDB; AAX30915.
 XX
 PT Nucleic acids encoding pneumococcal polypeptide(s) - useful in
 PT vaccines, drug screening, etc
 XX
 PS Claim 6; Page 325; 354pp; English.
 XX
 CC AAX30724 to AAX30946 represent genomic DNA sequences isolated from
 CC Streptococcus pneumoniae strain 0100993. These genomic DNA sequences
 CC encode the novel proteins given in AAY1114 to AAY11367. The proteins,
 CC isolated from Streptococcus pneumoniae, can be used in vaccines against
 CC streptococcal infections and in assays for identifying compounds that
 CC inhibit or activate the activity of the proteins. The antagonists can
 CC be used to treat an individual having need to inhibit a bacterial
 CC protein. Vectors expressing the proteins can be used to induce a
 CC protective immune response in mammals.
 XX
 SQ Sequence 171 AA;

Query Match 33.7%; Score 47.5; DB 18; Length 171;
 Best Local Similarity 34.6%; Pred. No. 6.3;
 Matches 9; Conservative 9; Mismatches 7; Indels 1; Gaps 1;

QY 3 CVLGIHIVPEFIESQLGLSPVSL 28
 |:|:|:|:|:|:|:|:|:|:
 Db 90 clyghlhpvsawlgk1-flnpgsl 114

RESULT 6
 ID AAW38569 standard; Protein; 173 AA.
 XX
 AC AAW38569;
 XX
 DT 06-NOV-1998 (first entry)
 XX
 DE Streptococcus pneumoniae protein of unknown function.
 XX
 KM Streptococcus pneumoniae protein; genetic immunisation; antagonist;
 KM immunological response; inoculation; antibody production; inhibitor;
 KM T cell immune response; antimicrobial compound; bacterial adhesion;
 KM extracellular matrix protein; protein-mediated cell invasion; wound;
 KM pathogenesis.
 XX
 OS Streptococcus pneumoniae.
 XX
 XX WO9743303-A1.
 XX
 PD 20-NOV-1997.
 XX
 PF 14-MAY-1997; 97WO-US07950.
 XX
 PR 14-MAY-1996; 96US-0017670.
 XX
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 XX (SMIK) SMITHKLINE BEECHAM PLC.
 XX
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
 PI Stodola RK;
 XX
 DR WPI: 1998-008793/01.

CC as sequence-determined DNA fragments (SDFs), from corn plants and
CC Arabidopsis thaliana. The SDFs are promoters, structural genes,
CC untranslated regions (UTRs), or 3' termination sequences. They can be
CC used for expressing a gene product and controlling expression of a
CC target gene, either as a promoter, a structural gene, an UTR or as a
CC 3' termination sequence. They are also useful as tools for genetic
CC mapping, and identification of a particular individual plant or for
CC clustering a group of plants with a common trait. AAA78433 to AAA78630
CC and AAB24605 to AAB25099 represent the specifically claimed
CC polynucleotide sequences and polypeptides encoded by them given in the
CC present invention.

XX
XX
SQ Sequence 127 AA;

Query Match 32.6%; Score 46; DB 21; Length 127;
Best Local Similarity 36.8%; Pred. No. 7.9;
Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

OY 9 HIVEPIESQLGLSPVS 27
||||::: ||: ||
Db 51 hivpqwidfsvlsimpfs 69

RESULT 9

AAB24843
ID AAB24843 standard; Peptide: 169 AA.

XX
XX AAB24843;

DT 27-NOV-2000 (first entry)

XX Plant SDF encoded polypeptide sequence SEQ List 1 NO:338.

DE Plant; corn; Arabidopsis thaliana; sequence-determined DNA fragment;
KM SDF; genetic mapping; identification; promoter; structural gene; UTR;
KW untranslated region; expression control.

XX
XX Plant.

OS
XX
XX WO200040695-A2.

PN 13-JUL-2000.

PD 07-JAN-2000; 2000WO-US00466.

PF 08-JAN-1999; 99US-0115293.

PR (CERE-) CERES INC.

XX
XX
XX Alexandrov N, Brover V, Chen X, Subramanian G, Troukhan ME;
PI Zheng L;

PI WPI; 2000-465970/40.

XX
XX
XX New corn plant and Arabidopsis thaliana sequence-determined DNA
PT fragments, useful for expressing gene products and for controlling
PT expression of a target gene -

XX
XX
XX Claim 14; Page 485-486; 673pp; English.

XX The present invention describes polynucleotides, such as complete cDNA
CC sequences and/or sequences of genomic DNA encompassing complete genes,
CC portions of genes, and/or intergenic regions, collectively referred to
CC as sequence-determined DNA fragments (SDFs), from corn plants and
CC Arabidopsis thaliana. The SDFs are promoters, structural genes,
CC untranslated regions (UTRs), or 3' termination sequences. They can be
CC used for expressing a gene product and controlling expression of a
CC target gene, either as a promoter, a structural gene, an UTR or as a
CC 3' termination sequence. They are also useful as tools for genetic
CC mapping, and identification of a particular individual plant or for
CC clustering a group of plants with a common trait. AAA78433 to AAA78630
CC and AAB24605 to AAB25099 represent the specifically claimed

CC polynucleotide sequences and polypeptides encoded by them given in the
CC present invention.

XX
XX
SQ Sequence 169 AA;

Query Match 32.6%; Score 46; DB 21; Length 169;
Best Local Similarity 36.8%; Pred. No. 11;
Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

OY 9 HIVEPIESQLGLSPVS 27
||||::: ||: ||
Db 93 hivpqwidfsvlsimpfs 111

RESULT 10

AAB24842
ID AAB24842 standard; Peptide: 181 AA.

XX
XX AAB24842;

DT 27-NOV-2000 (first entry)

XX Plant SDF encoded polypeptide sequence SEQ List 1 NO:337.

DE Plant; corn; Arabidopsis thaliana; sequence-determined DNA fragment;
KM SDF; genetic mapping; identification; promoter; structural gene; UTR;
KW untranslated region; expression control.

XX
XX Plant.

OS
XX
XX WO200040695-A2.

PN 13-JUL-2000.

PD 07-JAN-2000; 2000WO-US00466.

PF 08-JAN-1999; 99US-0115293.

PR (CERE-) CERES INC.

XX
XX
XX Alexandrov N, Brover V, Chen X, Subramanian G, Troukhan ME;
PI Zheng L;

PI WPI; 2000-465970/40.

XX
XX
XX New corn plant and Arabidopsis thaliana sequence-determined DNA
PT fragments, useful for expressing gene products and for controlling
PT expression of a target gene -

XX
XX
XX Claim 14; Page 485; 673pp; English.

XX The present invention describes polynucleotides, such as complete cDNA
CC sequences and/or sequences of genomic DNA encompassing complete genes,
CC portions of genes, and/or intergenic regions, collectively referred to
CC as sequence-determined DNA fragments (SDFs), from corn plants and
CC Arabidopsis thaliana. The SDFs are promoters, structural genes,
CC untranslated regions (UTRs), or 3' termination sequences. They can be
CC used for expressing a gene product and controlling expression of a
CC target gene, either as a promoter, a structural gene, an UTR or as a
CC 3' termination sequence. They are also useful as tools for genetic
CC mapping, and identification of a particular individual plant or for
CC clustering a group of plants with a common trait. AAA78433 to AAA78630
CC and AAB24605 to AAB25099 represent the specifically claimed
CC polynucleotide sequences and polypeptides encoded by them given in the
CC present invention.

XX
XX
SQ Sequence 181 AA;

Query Match 32.6%; Score 46; DB 21; Length 181;
Best Local Similarity 36.8%; Pred. No. 12;
Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 9 HIVPEFIESQLGLISPV 27
 |||:::|:|:|
 Db 105 hlvpqldfsvslmmpfs 123

RESULT 11

AAB90659
 ID AAB90659 standard; Protein; 183 AA.

AC AAB90659;

DE 01-JUN-2001 (first entry)

XX Bufo marinus lipocalin protein, SEQ ID NO: 202.

XX Toad; lipocalin; secreted protein; immunomodulatory; antisclerotic;
 KW dermatological; anti-inflammatory; anti-HIV; cytostatic; cardiant;
 KW vascular; anti-angiogenic; ophthalmological; neuroprotectant;
 KW nootropic; anticonvulsant; antialzheimer's; antiparkinsonian;
 KW antimicrobial; vulnery; vaccine; gene therapy; cancer;
 KW protein coordinate data; infection.

XX Bufo marinus.

XX WO200121658-A1.

XX 29-MAR-2001.

XX 22-SEP-2000; 2000WO-US26013.

XX 24-SEP-1999; 99US-0155709.

XX (HUMA-) HUMAN GENOME SCI INC.

XX NI J, Baker RP, Birse CE, Edner R, Piscella M, Komatsoulis GA;
 PI Lafleur DM, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;
 PI Young PR, Wel P, Florence KA;
 XX WPI; 2001-235311/24.

XX Nucleic acids encoding 32 human secreted polypeptides, useful for
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
 PT disease and diabetic retinopathy -

XX Disclosure; Fig 2; 890pp; English.

XX The present sequence is provided in a specification relating to nucleic
 CC acid molecules encoding 32 novel human secreted polypeptides. The nucleic
 CC acid molecules and polypeptides may be used in the prevention, diagnosis
 CC and treatment of diseases such as immune disorders (e.g. multiple
 CC sclerosis, systemic lupus erythematosus and human immunodeficiency virus
 CC (HIV) infections), hyperproliferative disorders (e.g. cancers and
 CC Gaucher's disease), cardiovascular diseases (e.g. Schmitz syndrome,
 CC Chaga's cardiomyopathy and coronary arteriosclerosis), angiogenic
 CC disorders (e.g. corneal graft neovascularisation and diabetic
 CC retinopathy), neurological disorders (e.g. Huntington's chorea,
 CC Alzheimer's disease and Parkinson's disease), infectious diseases and/or
 CC for promoting wound healing, regeneration and/or chemotaxis. The nucleic
 CC acid molecules may be used to produce the secreted polypeptides. They may
 CC also be used as DNA probes in diagnostic assays to detect and quantitate
 CC the presence of similar nucleic acid sequences in samples. The
 CC polypeptides may be used as antigens in the production of antibodies and
 CC in assays to identify modulators of their expression and activity.

XX Sequence 183 AA;

Query Match 32.6%; Score 46; DB 22; Length 183;

Best Local Similarity 47.4%; Pred. No. 12;
 Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 3 CVLGYLHIVPEFIESQLG 21

Db 17 cvygdvripqpdfgedk1lg 35
 |||:|:|:|:|:|

RESULT 12

AAW28030
 ID AAW28030 standard; Protein; 125 AA.

XX AAW28030;

DE 27-AUG-1998 (first entry)

XX Staphylococcus aureus protein of unknown function.

XX Staphylococcus aureus protein; ribozyme; antisense sequence; control;
 KW Staphylococcal gene; regulatory element; bacterial gene expression;
 KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
 KW toxic shock syndrome.

XX Staphylococcus aureus.

XX WO9730070-A1.

XX 21-AUG-1997.

XX 19-FEB-1997; 97WO-US02318.

XX 20-FEB-1996; 96US-0011888.

XX (SMIR) SMITHKLINE BEECHAM CORP.

XX Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;
 PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;

XX WPI; 1997-42469/39.

XX N-PSDB; AAT83984.

XX Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
 PT to isolate antimicrobial compounds, and in vaccines against S.
 PT aureus infection

XX Claim 6; Page 419; 989pp; English.

XX The present sequence represents a Staphylococcus aureus protein of
 CC unknown function. The DNA sequence was isolated from a library of
 CC clones of S. aureus WCUH 29 in Escherichia coli. The DNA sequence can
 CC be used in the construction of ribozymes and antisense sequences to
 CC control the expression of Staphylococcal genes. The DNA sequence is
 CC also useful as a source of regulatory elements for the control of
 CC bacterial gene expression. The present protein may be used to produce
 CC vaccines to enable a host to produce specific antibodies with
 CC antibacterial action. These vaccines and antibodies would protect
 CC a host against invasion by S. aureus, and conditions relating to
 CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
 CC skin syndrome, and toxic shock syndrome.

XX Sequence 125 AA;

Query Match 32.3%; Score 45.5; DB 18; Length 125;
 Best Local Similarity 30.3%; Pred. No. 9.3;
 Matches 10; Conservative 7; Mismatches 5; Indels 11; Gaps 1;

QY 5 LGYL-----HIVPEFIESQLGLISPV 26

Db 26 lgylveildrklhagfifefigsivglavl 58
 |||||:|:|:|:|:|

RESULT 13

AAAG24612
 ID AAG24612 standard; Protein; 296 AA.

XX AAG24612;

XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 28355.
DE
XX
KM Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130049.
PR 23-APR-1999; 990S-0130510.
PR 23-APR-1999; 990S-0130891.
PR 28-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 06-MAY-1999; 990S-0132487.
PR 07-MAY-1999; 990S-0132863.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139452.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.

PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139889.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144884.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145513.
PR 27-JUL-1999; 990S-0145518.
PR 27-JUL-1999; 990S-0145519.
PR 28-JUL-1999; 990S-0145591.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.

PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151308.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 31.9%; Score 45; DB 21; Length 296;
Best Local Similarity 63.6%; Pred. No. 30;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 7 YLHIVEPFES 17
Db 236 ylhivpfen 246

RESULT 14

AAG24611
ID AAG24611 standard; Protein; 328 AA.

XX AAG24611;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 28354.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.
XX Arabidopsis thaliana.
OS EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144684.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 04-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147302.
PR 06-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 23-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.

PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 22-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 23-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161992.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

Query Match 31.98; Score 45; DB 21; Length 328;
Best Local Similarity 63.68; Pred. No. 34;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 7 YLHVPERIES 17
||| |::|:
Db 268 ylhfpqfien 278

RESULT 15

ID MAG24610 standard; Protein; 354 AA.

AC MAG24610;
DT 17-OCT-2000 (first entry)
XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 28353.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX PD 06-SEP-2000.

XX	25-FEB-2000;	2000EP-0301439.	PR	08-JUL-1999;	99US-0142803.
PF			PR	09-JUL-1999;	99US-0142920.
XX	25-FEB-1999;	99US-0121825.	PR	12-JUL-1999;	99US-0142977.
PR	05-MAR-1999;	99US-0123180.	PR	13-JUL-1999;	99US-0143542.
PR	09-MAR-1999;	99US-0123548.	PR	14-JUL-1999;	99US-0143624.
PR	23-MAR-1999;	99US-0125788.	PR	15-JUL-1999;	99US-0144005.
PR	25-MAR-1999;	99US-0126264.	PR	16-JUL-1999;	99US-0144085.
PR	29-MAR-1999;	99US-0126785.	PR	19-JUL-1999;	99US-0144086.
PR	01-APR-1999;	99US-0127462.	PR	19-JUL-1999;	99US-0144325.
PR	06-APR-1999;	99US-0128234.	PR	19-JUL-1999;	99US-0144331.
PR	08-APR-1999;	99US-0128714.	PR	19-JUL-1999;	99US-0144332.
PR	16-APR-1999;	99US-0129845.	PR	19-JUL-1999;	99US-0144333.
PR	19-APR-1999;	99US-0130077.	PR	19-JUL-1999;	99US-0144334.
PR	21-APR-1999;	99US-0130449.	PR	20-JUL-1999;	99US-0144335.
PR	23-APR-1999;	99US-0130510.	PR	20-JUL-1999;	99US-0144352.
PR	23-APR-1999;	99US-0130891.	PR	20-JUL-1999;	99US-0144632.
PR	28-APR-1999;	99US-0131449.	PR	21-JUL-1999;	99US-0144884.
PR	30-APR-1999;	99US-0132048.	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	99US-0132407.	PR	21-JUL-1999;	99US-0145086.
PR	04-MAY-1999;	99US-0132484.	PR	22-JUL-1999;	99US-0145088.
PR	05-MAY-1999;	99US-0132485.	PR	22-JUL-1999;	99US-0145085.
PR	06-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-0145087.
PR	06-MAY-1999;	99US-0132487.	PR	22-JUL-1999;	99US-0145089.
PR	07-MAY-1999;	99US-0132863.	PR	22-JUL-1999;	99US-0145192.
PR	11-MAY-1999;	99US-0134256.	PR	23-JUL-1999;	99US-0145145.
PR	14-MAY-1999;	99US-0134218.	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134219.	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134221.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145911.
PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145912.
PR	19-MAY-1999;	99US-0134941.	PR	28-JUL-1999;	99US-0145919.
PR	20-MAY-1999;	99US-0135124.	PR	02-AUG-1999;	99US-0145951.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146386.
PR	24-MAY-1999;	99US-0135629.	PR	02-AUG-1999;	99US-0146388.
PR	25-MAY-1999;	99US-0136021.	PR	02-AUG-1999;	99US-0146389.
PR	27-MAY-1999;	99US-0136392.	PR	03-AUG-1999;	99US-0147038.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147204.
PR	01-JUN-1999;	99US-0137222.	PR	05-AUG-1999;	99US-0147302.
PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147192.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147260.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147303.
PR	08-JUN-1999;	99US-0138094.	PR	06-AUG-1999;	99US-0147416.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138847.	PR	09-AUG-1999;	99US-0147935.
PR	14-JUN-1999;	99US-0139119.	PR	10-AUG-1999;	99US-0148171.
PR	16-JUN-1999;	99US-0139452.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139453.	PR	12-AUG-1999;	99US-0148341.
PR	17-JUN-1999;	99US-0139452.	PR	13-AUG-1999;	99US-0148565.
PR	18-JUN-1999;	99US-0139454.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139455.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139456.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139457.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139460.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139462.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139463.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139750.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151065.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151066.
PR	22-JUN-1999;	99US-0139889.	PR	27-AUG-1999;	99US-0151080.
PR	23-JUN-1999;	99US-0140353.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140354.	PR	30-AUG-1999;	99US-0151303.
PR	24-JUN-1999;	99US-0140695.	PR	31-AUG-1999;	99US-0151438.
PR	28-JUN-1999;	99US-0140823.	PR	01-SEP-1999;	99US-0151930.
PR	29-JUN-1999;	99US-0140921.	PR	07-SEP-1999;	99US-0152363.
PR	30-JUN-1999;	99US-0141287.	PR	10-SEP-1999;	99US-0153070.
PR	01-JUL-1999;	99US-0141847.	PR	13-SEP-1999;	99US-0153758.
PR	01-JUL-1999;	99US-0142154.	PR	15-SEP-1999;	99US-0154018.
PR	02-JUL-1999;	99US-0142053.	PR	16-SEP-1999;	99US-0154039.
PR	06-JUL-1999;	99US-0142390.	PR	20-SEP-1999;	99US-0154779.
			PR	22-SEP-1999;	99US-0155139.
			PR	23-SEP-1999;	99US-0155486.

PR 24-SEP-1999; 99US-0156559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159285.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match

Best Local Similarity 31.98; Score 45; DB 21; Length 354;
Matches 7; Conservative 63.68; Pred. NO. 37;

Mismatches 3; Mismatches 1; Indels 0; Gaps 0;

OY 7 YLHIVPEFIES 17

Db 294 YLHIVPEFIES 304

Search completed: August 7, 2001, 13:28:36
Job time: 351 sec

This Page Blank (uspto)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ BLACK BORDERS

☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

☐ FADED TEXT OR DRAWING

☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING

☐ SKEWED/SLANTED IMAGES

☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS

☐ GRAY SCALE DOCUMENTS

☐ LINES OR MARKS ON ORIGINAL DOCUMENT

☒ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

This Page Blank (uspto)